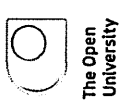
Non-Canonical cis-Acting Elements Regulating Polyadenylation of the Beta-Adducin Pre-mRNA

Thesis

UNRESTRICTED

Non-canonical cis-acting elements regulating polyadenylation of the beta-adducin pre-mRNA

Mirjana Nedeljković



A Thesis Submitted in Fulfillment of the Requirements of the Open University (UK)
for the Degree of Doctor of Philosophy
at the International Centre for Genetic Engineering and Biotechnology (ICGEB) Trieste, Italy

Director of Studies Dr. Andrés F. Muro
External Supervisor Dr. André Furger

June 2012

Date of Submission: 12 June 2012
Date of Award: 28 August 2012

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

Chemical compounds are named according to IUPAC recommendations.

List of frequently used Abbreviations:

3'RACE	3' Rapid Amplification of cDNA Ends
3'UTR	3' Untranslated Region
ACE	Adenylation Control Element
ARE	Adenine and uridine Rich Element
bp	base pair
cDNA	complementary DNA
CF I	Cleavage factor I
CF II	Cleavage factor II
CPSF	Cleavage and Polyadenylation Specificity Factor
CPE	Cytoplasmic Polyadenylation Element
CPEB	CPE Binding-protein
CstF	Cleavage stimulatory Factor
CTD	Carboxyl - terminal domain
DSE	Downstream Sequence Element
ESE	Exonic Splicing Enhancer
EST	Expressed Sequence Tag
Hm	Hexanucleotide motif
hnRNP	Heterogeneous nuclear ribonucleoproteins
Kb	Kilo base
mRNA	messenger RNA
nt	nucleotides
PABP	PolyA binding protein
PAP	PolyA polymerase
PAS	PolyAdenylation Site
PCR	Polymerase Chain Reaction
qPCR	quantitative PCR
RNAPII	RNA Polymerase II
RRM	RNA Recognition Motif
RT	Reverse Transcription
SR	Arginine-serine rich protein
ss	Splice site
USE	Upstream sequence element

ABSTRACT

The beta adducin gene has a very complex architecture, with tissue-specific use of promoters and polyadenylation sites (PAS). Its expression is restricted to neuronal and hematopoietic tissues. In mice, usage of a distal brain-specific PAS (PAS4) generates an 8.3 kb long mRNA with an unusually long 3'UTR of about 5.7 kb. Instead, the use of proximal erythroid-specific PASs results in short 3.1-3.7 kb mRNA isoforms. The presence and use of the different alternative PASs in the beta adducin gene are very well conserved among species suggesting an important and specific role for each of the generated 3'UTRs.

To study the regulatory mechanisms involved in beta adducin polyadenylation, minigene constructs containing the mouse beta adducin polyadenylation signals were used. The constructs were transfected into HeLa cells and their expression and PAS selection determined by Northern blot analysis. The exclusive usage of the PAS4 in HeLa cells has been observed. All the elements defining the core polyadenylation signal were characterized: the hexanucleotide motif (Hm), the upstream and downstream sequence elements (USE and DSE, respectively) elements, and the cleavage site.

Interestingly, I have detected the presence of two novel non-canonical cis-acting elements regulating 3'end processing at the PAS4. Both elements act on long-distance and to the best of our knowledge, long-distance upstream polyadenylation regulatory elements have not previously been described for the non-viral eukaryotic transcripts.

The first of these elements was essential to enable polyadenylation at the PAS4. It was located in a region spanning 355 nucleotides and includes the stop codon of beta adducin. The second non-canonical upstream polyadenylation regulatory element seems to inhibit processing at the PAS4. It was located in the region close to the second proximal PAS, about 4.5 kb upstream of the PAS4.

These results highlight the complexity of the regulatory mechanisms directing beta adducin pre-mRNA 3'end processing.

Chapter 1

Introduction

1.1 Pre-mRNA processing

Eukaryotic protein-coding gene expression is a very complex and tightly regulated process. One of the reasons for the complexity of the eukaryotic protein-coding gene expression lies in the regulation of the processing of the primary transcripts (pre-mRNAs). Pre-mRNA processing includes 5'end capping, editing, splicing and 3'end processing events. Once processed in the nucleus, mature mRNAs are exported to the cytoplasm and translated.

1.1.1 5'end capping

When the nascent transcript reaches a length of 25-30 nucleotides, the 7-methyl-guanylate cap structure is added to its 5' end. The 5'cap structure is subsequently recognized by the cap binding protein complex (CBC). This complex protects the nascent transcript from 5'-3' exonuclease degradation. It is also important for pre-mRNA splicing, 3'end processing and nuclear export of the mature mRNA and its translation (Lewis and Izaurralde, 1997; Sonenberg, 1988).

1.1.2 Editing

RNA editing is a rare form of RNA processing in which the nucleotide sequence of the primary transcript is altered from that encoded by the genome (Keegan et al., 2001). **C**ytosine (**C**) to **U**racil (**U**) and **A**denosine (**A**) to **I**nosine (**I**) editing have been described for mammals. Site-specific **C** to **U** editing in the apolipoprotein B pre-mRNA introduces a stop codon resulting in the synthesis of a truncated protein, apoB48. **A** to **I** editing has been found to alter the amino-acid composition of the ionotropic glutamate receptor subunit GluR-B leading to a decrease in Ca2+ permeability. Transcripts that undergo **A** to **I** editing are mostly expressed in the brain. Aberrant editing has been associated with certain cancers and a range of abnormal behaviors including epilepsy and depression (Brusa et al., 1995; Gurevich et al., 2002; Maas et al., 2001).

1.1.3 Splicing

Before being transported to the cytoplasm where the mature transcripts can be translated, intronic sequences are removed from the pre-mRNAs and exons are joined together in a process called "splicing". Splicing, which essentially is a two-step transesterification reaction is carried out by a multicomponent complex, the spliceosome. The spliceosome is a dynamic ribonucleoprotein complex composed of five snRNAs (U1, U2, U4, U5 and U6) and numerous associated proteins (Jurica and Moore, 2003). Components of the spliceosome interact with defined sequences at the exon/intron boundaries, the so called splice sites (5' and 3' splice site, respective to the intron), the polypyrimidine tract and the branch point sequence. Upstream from the typical **AG** dinucleotide at the 3' splice site a polypyrimidine tract (PPT) is present. The branch point sequence (BPS) is located upstream from the PPT. In higher eukaryotes it does not have a clear consensus sequence, exept for the presence of an **A**denosine which serves as the branch point in the splicing reaction. Splice sites are defined by **GU** and **AG** dinucleotides positioned at the 5' and 3' ends of the introns, respectively. However, other sequences flanking these dinucleotides are necessary, which help to define

the strength of the splice sites. The 5' splice site has the consensus **CAG:GURAGU** (where : denotes the exon-intron borders) and deviation from this consensus affects the strength of the splice site. In the same way, the presence of a polypyrimidine tract interrupted by purines weakens the 3' splice site. Exons that contain weak core splice sites need additional auxiliary elements to procure efficient splicing, named exonic splicing enhancers (ESE) and/or intronic splicing enhancers (ISE). Moreover, ESEs may also be necessary for the correct exon recognition of exons separated by large introns . Many ESE sequences are characterized by the enrichment in purines, which provide binding sites for proteins of the SR protein family (Busch and Hertel, 2012). Each member of the SR protein family contains a characteristic arginine/serine rich domain (RS domain) and one or two RNA Recognition Motifs (RRM). There are twelve SR protein members detected in humans, named Serine/Arginine-rich Splicing Factors 1 to 12 (SRSF1-12) (Manley and Krainer, 2010). The most studied ones are SRSF1 and 2 (formerly known as ASF/SF2 and SC35, respectively). The ESE sequences recognized by these factors have degenerated purine-rich structure, but are nevertheless specific for each of them. In addition to the SR protein family, there is a group of SR-related proteins that can interact with ESEs. The SR-related proteins are also characterized by the presence of a RS domain, whereas the presence of the RRM is optional. Some of the members of this family are splicing factors which associate with the major spliceosomal components, such as U1snRNP (U170K factor), or with the polypyrimidine tract (U2AF35 and 65). Others act as splicing coactivators, for example the SRm160/300 complex, which lacks the RRM and acts in promoting purine-rich ESE function in splicing through its interaction with other factors that are directly bound to the ESE (Blencowe, 2000). Moreover, along the exons and introns different silencer elements can be found: exonic and intronic splicing silencers (ESS and ISS, respectively). These elements are recognized by a group of heterogeneous nuclear ribonucleoproteins (hnRNPs) which, apart from their role in splicing, have a wide range of cellular functions (Han et al., 2010). hnRNPs bind to the silencer sequences through their RNA binding domain, which is mostly represented by one to four RRMs. RNA binding motifs

are deviated and specific for each member of the hnRNP family. A well studied member of the hnRNP family, hnRNP I, widely known as PTB (Polypyrimidine Tract Binding protein), acts in the repression of exon or intron definition through the binding to CU-rich silencer sequences. hnRNPs F/H operate as splicing repressors recognizing G-rich sequences while hnRNP L binds CA repeats. Together, SR proteins and hnRNPs, through their binding to enhancer and silencer elements, allow fine tuning in the regulation of the pre-mRNA splicing and determine alternative splicing patterns. Moreover, SR proteins and hnRNPs are the main regulators of the alternative splicing events.

1.1.4 3'end processing

The final pre-mRNA processing reaction required for the maturation of pre-mRNAs is 3'end processing. 3' end processing is a two step reaction in which the pre-mRNA is first cleaved and subsequently polyadenylated. The only mRNAs that are not polyadenylated are the transcripts of the replication-dependent histone genes. These transcripts are matured at their 3'ends by a cleavage event only, directed by the U7snRNP (Dominski and Marzluff, 2007). As for the splicing reaction, specific cis-acting elements that are recognized by specific trans-acting factors are necessary for correct and efficient 3'end processing. The core elements that are involved in the correct definition of the place where the cleavage and polyadenylation will occur represent the polyadenylation signal. Mammalian polyadenylation signals are characterized by the presence of the hexanucleotide motif (**AAUAAA** or similar sequence located 20-30 bases upstream of the cleavage site), subsequently named "Hm", the **U**-rich and/or **U/GU**-rich elements located immediately upstream and downstream of the cleavage site, respectively (upstream sequence element or USE and downstream sequence element or DSE) and the cleavage site itself, named also poly(A) site (PAS) as it is the site of the polyA tail addition (Figure 1) (Proudfoot, 2011).

The 3'end processing machinery that recognizes and interacts with these elements of the mammalian polyadenylation signals is composed of several multisubunit protein

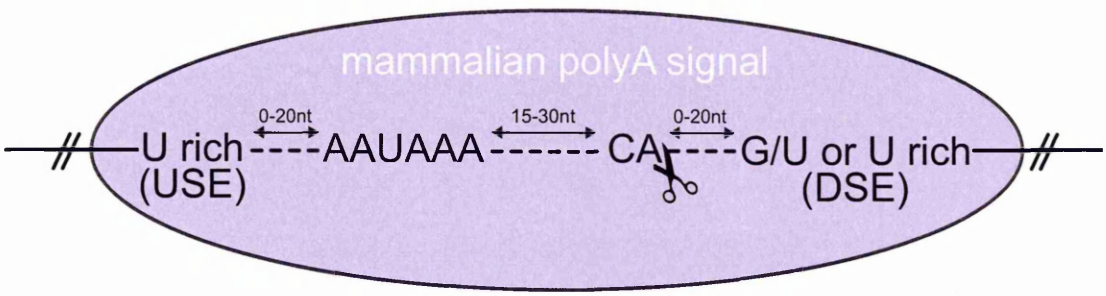


Figure 1
Schematic representation of the sequence composition of the mammalian polyadenylation signal.
USE - upstream sequence element; DSE – downstream sequence element; double arrowheads indicate the distance in nucleotides (nt) between the indicated elements; the scissors indicate the position of the cleavage site; adapted from Proudfoot 2011.

factors: Cleavage and Polyadenylation Specificity Factor (CPSF), Cleavage stimulatory Factor (CstF), Cleavage Factor I and II (CFI and CFII); as well as of the single polypeptides: PolyA Polymerase (PAP), PolyA Binding Proteins (PABPs), symplekin and the large subunit of the RNA Polymerase II (RNAP II). A recent study showed that the human 3'end processing complex is composed of at least ~85 proteins (Shi et al., 2009). In addition to the main cleavage and polyadenylation factors, more than 50 of the protein factors identified link 3'end processing of pre-mRNAs with other cellular processes.

1.1.4.1 The 3'end processing complex assembly

The assembly of the mammalian 3'end processing complex is shown in Figure 2 and is described below. The main characteristic of the 3'end processing complex assembly is the recruitment of CPSF and CstF onto sequences present upstream and downstream of the cleavage site, the Hm and the DSE, respectively.

Components of CPSF are CPSF-160, CPSF-100, CPSF-73, CPSF-30, hFip1 and WDR33. The **AAUAAA** hexanucleotide motif is recognized and bound by the CPSF-160 subunit (Keller et al., 1991; Murthy and Manley, 1995). CPSF-73 and CPSF-100 subunits, together with symplekin, form the core cleavage complex (Sullivan et al., 2009). This complex is assembled at the cleavage site. CPSF-30 and hFip1 are bound to the **U**-rich sequences found immediately upstream of the cleavage site (Barabino et al., 1997; Kaufmann et al., 2004). WDR33, a recently identified component of CPSF, is in close connection with CPSF-73 (Shi et al., 2009).

CstF is composed of three subunits, CstF-50, 64 and 77, each of them likely to form homodimers (Moreno-Morcillo et al., 2011). CstF-50 binds to the RNAP II carboxi-terminal domain (RNAP II CTD). CstF-64 is responsible for the recognition and binding to the **U/GU**-rich DSE (MacDonald et al., 1994). CstF-77 is important for the interaction with the CPSF-160 subunit, and it is also mediating the binding of the other subunits of CstF.
CFI is a heterotetrameric protein composed of the CFI-25 homodimer linked to CFI-59 or CFI-68 homodimers. Through the CFI-25 subunit, CFI recognizes **UGUA** elements present in the 3'UTR (Brown and Gilmartin, 2003; Yang et al., 2010).
CFII is believed to associate transiently with the 3'end processing complex, as one of its subunits, Clp1, is completely absent and the other one, Pcf11, is present only at low levels in the recently analyzed human 3'end processing complex (Shi et al., 2009). However, its role is of high importance in the stimulation of cleavage, as a purified human 3'end processing complex was inactive in cleavage until the addition of CFII and CFI (Shi et al., 2009).
PAP is anchored to the 3'end processing complex by CPSF-160 and hFip1 (Kaufmann et al., 2004). CPSF-73 is the endonuclease that cleaves the pre-mRNA (Mandel 2006, Nature), after which PAP catalyzes the nontemplated addition of adenosines. At this point the CstF, CFI and CFII dissociate from the polyadenylation complex. Afterwards, nuclear PABP is recruited on the growing poly**A** tail, which enables PAP to proceed rapidly until the addition of ~250 adenosine residues (Figure 2B). The serine/threonine phospatase PPI has been recently described to be an additional factor important for the polyadenylation step (Shi et al., 2009). The remaining fragment downstream of the cleavage site is rapidly degraded by the nuclear 5'-3' exonuclease (Xrn2) (Figure 2B). Xrn2 stimulates transcription termination, by a so far unknown mechanism.

1.1.4.2 The hexanucleotide motif (Hm)

Observing the high variability in sequence composition of the polydenylation signal elements, it is quite fascinating how the 3'end processing factors recognize the

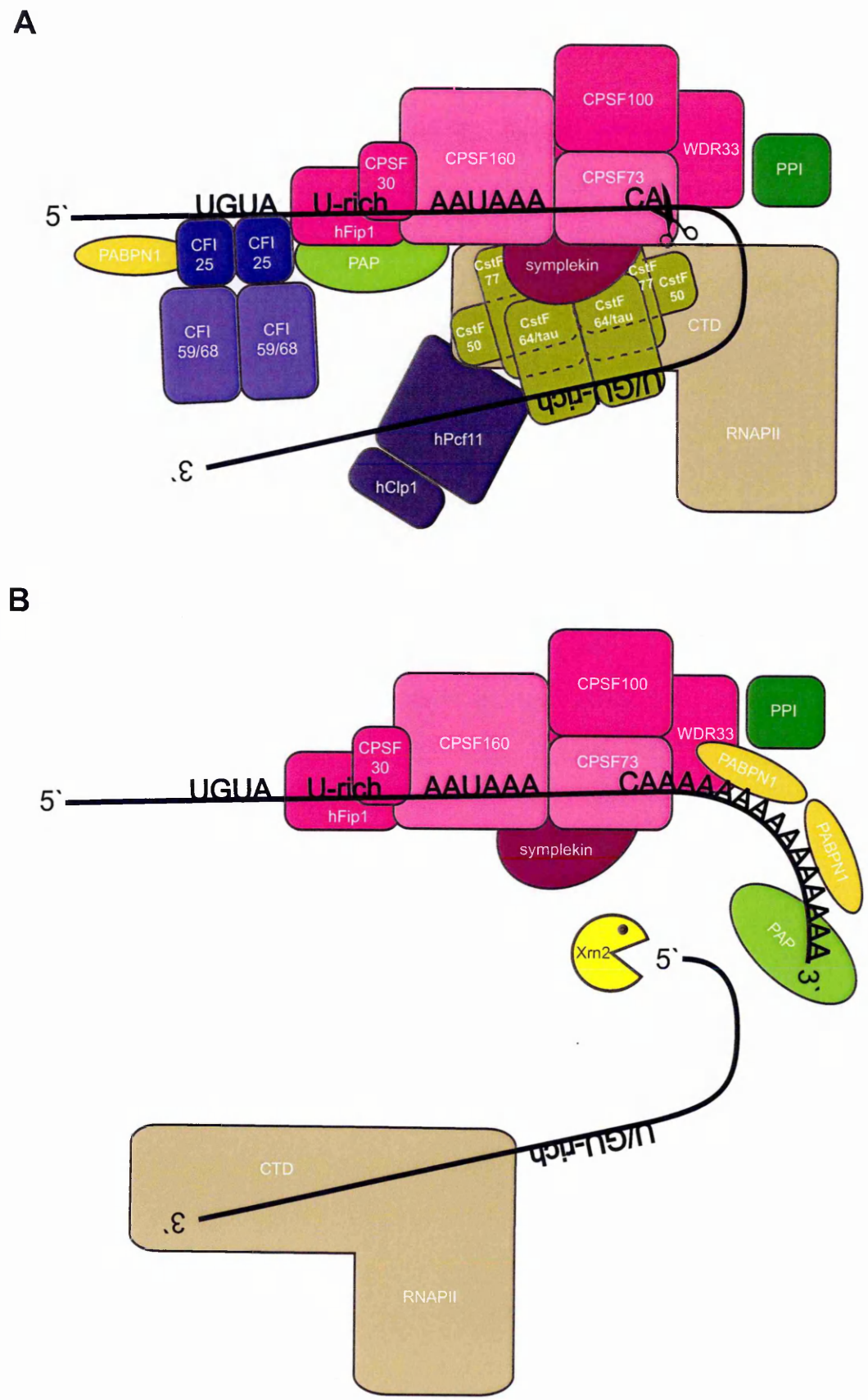


Figure 2
Mammalian 3'end processing complex assembly.
For detailed explanation see text.

polyadenylation signals. The only element that has a well-defined consensus sequence is the highly conserved hexanucleotide motif, **AAUAAA**. Thus, it is not surprising that the **AAUAAA** was the first polyadenylation signal element to be described (Proudfoot and Brownlee, 1976). *In vitro* studies in which point mutations in the **AAUAAA** were introduced demonstrated the specificity and importance of this element for efficient cleavage and polyadenylation of pre-mRNAs (Sheets et al., 1990). All the introduced point mutations led to the almost complete abolishment of polyadenylation, apart from the mutation of the second **A** to **U** (the **AUUAAA** hexanucleotide), which showed the ~80% of the wild type hexanucleotide activity (Sheets et al., 1990)(Figure 3). This study mirrors the situation of thalassemias where different mutations in the **AAUAAA** sequence of the alpha or beta globin gene affect globin mRNA levels, resulting in abnormal hemoglobin production (Chen et al., 2006). Availability of the whole genome sequences, cDNAs and expressed sequence tags (ESTs) data allowed the bioinformatic analyses of the most representative hexanuleotide motifs upstream of the cleavage site in a genome wide context. These analyses confirmed the importance of the **AAUAAA** sequence, showing the presence of this Hm in ~60% of human and mouse genes (Beaudoing et al., 2000; Tian et al., 2005) (Figure 3). The counterpart of the **AAUAAA** sequence, the **AUUAAA** sequence is present in ~15% of the analyzed genes. Nowadays, both the **AAUAAA** and the **AUUAAA** hexamers are considered as canonical ones. All the other detected hexanucleotide motifs, the non-canonical ones, are present in about 20-30% of the analyzed genes and most of them contain a single base difference in respect to the **AAUAAA** (Figure 3). For about 8% of the analyzed genes no clear hexamer can be assigned (Tian et al., 2005).

1.1.4.3 The cleavage (polyA) site

Initial analysis of the pre-mRNA cleavage process showed that the cleavage site is mostly represented by a **CA** dinucleotide (Sheets et al., 1990). Further studies demonstrated that the order of the preferred nucleotide at the cleavage site is **A>U>C>>G** (Chen

hexamer variant	frequency(%) Hs	Mm	polyA efficiency(%)
AAUAAA	53.18	59.16	100
AUUAAA	16.78	16.11	77 ± 4.7
UAUAAA	4.37	3.79	17 ± 3.0
AGUAAA	3.72	3.28	29 ± 8.1
AAGAAA	2.99	2.15	6 ±1.0
AAUAUA	2.13	1.71	10 ± 2.3
AAUACA	2.03	1.65	11 ± 2.3
CAUAAA	1.92	1.80	18 ± 6.4
GAUAAA	1.75	1.16	11 ± 1.0
AAUGAA	1.56	0.90	4.3 ± 0.6
UUUAAA	1.20	1.08	/
ACUAAA	0.93	0.64	11 ± 6.0
AAUAGA	0.60	0.36	3.3 ±1.5

Figure 3
Correlation between the frequency and polyadenylation efficiency of different hexanucleotide motif variants.
Data regarding frequency are taken from Tian et al 2005.; Hs - Homo sapiens; Mm - Mus musculus; polyadenylation efficiency data are taken from Sheets et al 1991.

et al., 1995). Both of these studies demonstrated that the mutation of the nucleotide at the cleavage site position does not interfere with the cleavage efficiency, but shift the cleavage site to a close nucleotide. Thus, pre-mRNA cleavage is a rather imprecise process and does not necessarily occur at a specific nucleotide. This observation was confirmed by other studies, which demonstrated the presence of multiple cleavage sites for a single polyadenylation signal (Pauws et al., 2001; Tian et al., 2005). It has been estimated that 51.25% of human and 46.97% of mouse polyadenylation signals have more than one cleavage site (Tian et al., 2005). Although these results show high heterogeneity in the selection of the cleavage site, the site of cleavage is always placed at the certain distance in between the **AAUAAA** motif and the DSE (Chen et al., 1995) (Figure 1).

1.4.4.4 The downstream sequence element (DSE)

The DSE is usually represented by a number of **U**-rich or/and **G/U**-rich elements placed immediately downstream of the cleavage site (Zarudnaya et al., 2003) (Figure 1), but

there is no defined consensus sequence for the DSE. Unlike the **AAUAAA** motif, in which point mutations have a severe impact on the efficiency of 3'end processing, mutations of the DSE had a less pronounced influence on the 3'end processing of the tested pre-mRNAs, perhaps due to its divergence in sequence composition (Zarudnaya et al., 2003; Zhao et al., 1999). Studies of the DSEs were mostly conducted in the context of strong polyadenylation signals that are represented by the canonical **AAUAAA** Hm. Some of the well characterized DSE of viral and cellular polyadenylation signals are represented in Figure 4 (Zarudnaya et al., 2003).

The importance of the DSE for the recognition of the polyadenylation signal in the absence of the canonical Hm was highlighted in the study of the human intronless melanocortin 4 receptor gene (MC4R) 3'end processing (Nunes et al., 2010). The MC4R pre-mRNA cleavage efficiency depends mostly on the presence of a strong DSE, whereas the hexanucleotide motif appears to be less important. In fact, in the MC4R gene the natural hexamer can be mutated and an **A**-rich upstream element is sufficient, together with the strong MC4R DSE, to direct efficient 3'end processing. This observation was extended to the whole genome. A genome-wide bioinformatic analysis showed that non-canonical polyA signals characterized by the presence of A-rich upstream sequences are usually associated with stronger DSE, characterized by downstream sequences richer in **U-** and **G/U** than DSEs of genes having canonical Hm (Nunes et al., 2010). These experiments highlight the role of the DSE in the polyA site definition in the absence of the canonical Hm.

In addition, another sequence positioned just downstream of the core DSE, called the **G**-rich auxiliary DSEs, has been described to enhance 3'end processing of various pre-mRNAs (Zarudnaya et al., 2003). Interestingly, the **G**-rich element placed far downstream of the cleavage site (440 nucleotides downstream) has been reported to be important for the efficient 3'end processing of the melanocortin receptor 1 gene (MC1R) (Dalziel et al., 2007). Members of the hnRNP H protein family are bound to the G-rich auxiliary DSEs and stimulate 3'end processing probably by enhancing the

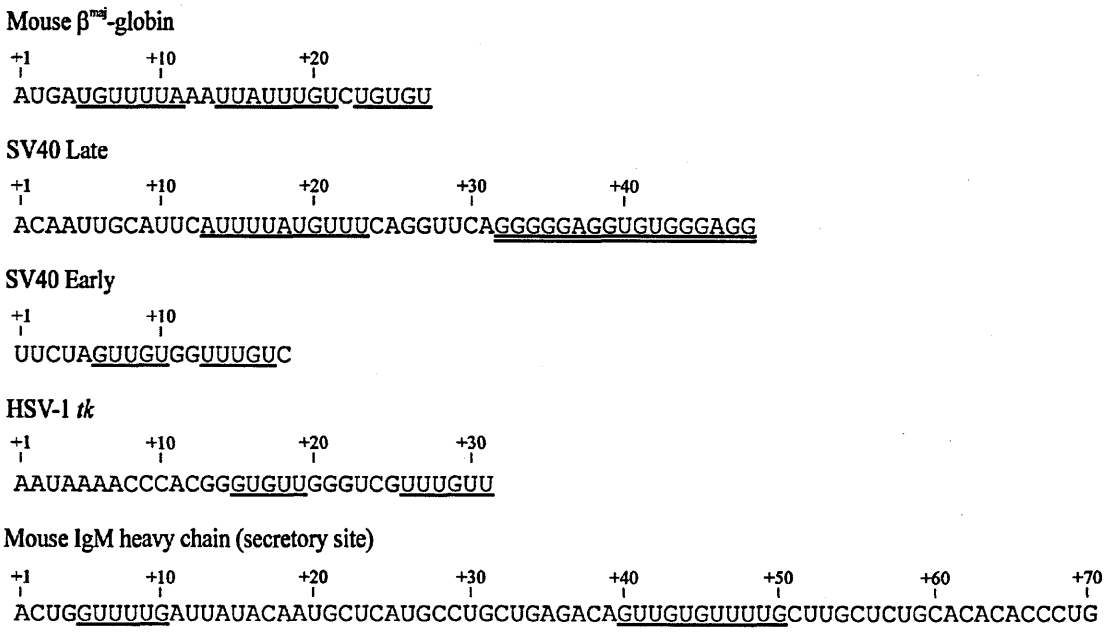


Figure 4
Sequence composition of downstream sequence elements found in viral and mammalian polyadenylation signals.
U- and GU-rich tracts are underlined; double line indicates G-rich tract; +1 indicates the first nucleotide after the cleavage site; adapted from Zarudnaya et al 2003.

efficiency of the CstF64 assembly on the core DSE (Arhin et al., 2002; Bagga et al., 1998; Dalziel et al., 2007).

1.1.4.5 The upstream sequence element (USE)

Additional elements enhancing 3'end processing can be found immediately upstream of the hexanucleotide motif: the upstream sequence elements (USEs) (Figure 1). Similar to the DSEs, the USEs have no defined consensus sequence motif and are generally represented by a variable number of **U**-rich elements. Bioinformatic analysis of the human and mouse genes showed a general **AU** enrichment in the 100 bases upstream of the cleavage site (Tian et al., 2005). Some of the well described USEs are represented in Figure 5 and are discussed below. The first studies of the USE were conducted on viral polyadenylation signals. The USE of the SV40 late polyadenylation signal is represented by the three **AUUUGURA** sequence motifs (Schek et al., 1992). This element is recognized by the U1 snRNP-A protein, which interacts with CPSF-160 stabilizing its binding

SV40 late

AUGCAGUGAAAAAAAUGCUUUAUUUGUGAAAUUUGUGAUGCUAUUGCUUUAUUUGUAACCAUUAUAAGCUGCAAUAAA

C2 complement

GACUUGACUCAUGCUUGUUUCACUUUCACAUGGAAUUUCCCAGUUAUGAAAUAAUAAA

lamin B2

AUUCGGUUUUUAAGAAGAUGCAUGCCUAACGUGUUCUUUUUUUUUUUCCAAUGAUUUGUAAUAUACAUUUUAUGACUGG

AAACUUUUUUGUACAACACUCCAAUAAA

papolG

GUUGUUUGUAAAUGAAGUCGUAUGUAUUUUCAGAGUAUUUUUGUAUGUACUGUAAGAUACCAUCUUUUCAAAGAGAAA

Figure 5
Sequence composition of upstream sequence elements found in viral and mammalian polyadenylation signals.
U- and GU-rich tracts are underlined; open box indicates hexanucleotide motif.

to the **AAUAAA**, and thus enhancing polyadenylation efficiency (Lutz and Alwine, 1994; Lutz et al., 1996). The first non-viral example of the USE was described for the human C2 complement gene (Moreira et al., 1995), which is bound by PTB enhancing the cleavage efficiency (Moreira et al., 1998). As for the SV40, the USE of the human lamin B2 polyadenylation signal enhances cleavage and polyadenylation by increasing the stability of the CPSF-**AAUAAA** complex, although by a different mechanism (Brackenridge and Proudfoot, 2000). The stabilization of the CPSF-**AAUAAA** complex can be achieved also through the direct binding of the CPSF components, CPSF-30 and hFip1, to the **U**-rich USEs (Barabino et al., 1997; Kaufmann et al., 2004). The importance of the USEs is highlighted in the recognition of the polyadenylation signals having the non-canonical hexanucleotide motif (Venkataraman et al., 2005). These USEs are characterized by the presence of the **UGUAN** consensus sequence and are recognized by the CFI. This interaction stimulates polyadenylation in the absence of the canonical **AAUAAA**, through the recruitment of the hFip1 and PAP to the RNA substrate.

The above mentioned studies introduced only a few concepts of the highly complex world of pre-mRNA 3'end processing. The strength of the polyadenylation signal depends on different combinatorial codes of its multiple sequence elements, which are recognized by a specific set of 3'end processing factors.

1.1.5 Crosstalk between the cotranscriptional pre-mRNA processing events

Pre-mRNA processing events take place mostly cotranscriptionaly (Proudfoot et al., 2002). They are in close connection with the RNAP II through its CTD, which serves as a landing pad for different processing factors. Moreover, the pre-mRNA processing events are also interconnected and influence one another.

1.1.5.1 3'end processing and transcription

Transcription and 3'end processing factors are interconnected from the initiation until the termination of the transcription. TFIID and TFPBII associate with CPSF and CstF, respectively, and bring them to the preinitiation complex (Dantonel et al., 1997; Wang et al., 2010). When transcription initiates, CPSF and CstF are recruited onto elongating RNAP II CTD and carried all the way until RNAP II transcribes polyadenylation signal. Then, CPSF and CstF are recruited from the RNAP II CTD onto the pre-mRNA polyadenylation signal. This recruitment is followed by the incorporation of the other 3'end processing factors completing complex assembly and leading to cleavage and polyadenylation of the nascent transcript. As previously mentioned, the CTD of RNAPII is an active component of the 3'end processing complex. 3'end processing has an important role in the termination of RNAP II mediated transcription (Buratowski, 2005; Proudfoot, 2004). As mentioned above, the remaining downstream cleaved RNA product is rapidly degraded by the 5'-3' exonuclease Xrn2 (Figure 2B). By the proposed torpedo model of trancription termination, it is believed that once Xrn2 reaches RNAP II, it causes dissociation of RNAP II from the DNA template, promoting termination of transcription (Buratowski, 2005). Moreover, additional elements, placed downstream of the polyadenylation site, have been described to influence transcription termination. Such elements are cotranscriptional cleavage sites (CoTC) and/or RNAP II transcriptional pause sites, usually represented by **G**-rich sequence (Dye and Proudfoot, 2001; Gromak et al., 2006). Additionally, all the assembled cleavage and polyadenylation factors remain on the pre-mRNA after the cleavage event. Polyadenylation and,

in the case of intron-containing pre-mRNAs, splicing events are necessary to release the transcript from the RNAP II polymerase (Rigo and Martinson, 2009).

A recent study demonstrated that 3'end processing facilitates recycling of RNAP II and other transcription factors to the promoter region after the first round of transcription, thus stimulating transcription initiation (Mapendano et al., 2010). This finding is in agreement with studies showing physical connections between promoter and polyadenylation site regions, termed gene loops (Perkins et al., 2008; O`Sullivan et al., 2004). Thus, the presence of gene loops could facilitate recycling of transcription factors from the 3'end of genes to the promoter.

1.1.5.2 3'end processing and 5'end capping

The pre-mRNA 5'end capping and 3'end processing interaction is established through direct connection of the nuclear cap binding complex (CBC) and 3'end processing complex (Flaherty et al., 1997). This interaction maintains the loop structure of the pre-mRNA and mature mRNA necessary for its stability and nuclear export. Moreover, the interaction of the CBC and 3'end processing complex is important for the efficient cleavage reaction.

1.1.5.3 3'end processing and splicing

The crosstalk between pre-mRNA 3'end processing and splicing reactions is emphasized in terminal exon definition and last intron removal (Martison 2011) (Figure 6). The terminal exon is a very particular case since it has only the 3'splice site, whereas the 5'splice site is lacking. Terminal exon definition depends on the interaction between splicing factors recognizing its 3'splice site and 3'end processing factors bound to the polyadenylation signal core elements. Early *in vitro* studies showed that the presence of an upstream intron in the pre-mRNA leads to enhancement of 3'end processing (Niwa et al., 1990). Additionally, the polyadenylation signal can also influence splicing of proximal introns (Niwa et al., 1991). *In vitro* studies showed that PAP is able to interact with the splicing factor U2AF65 enhancing its binding to the polypyrimidine tract of the

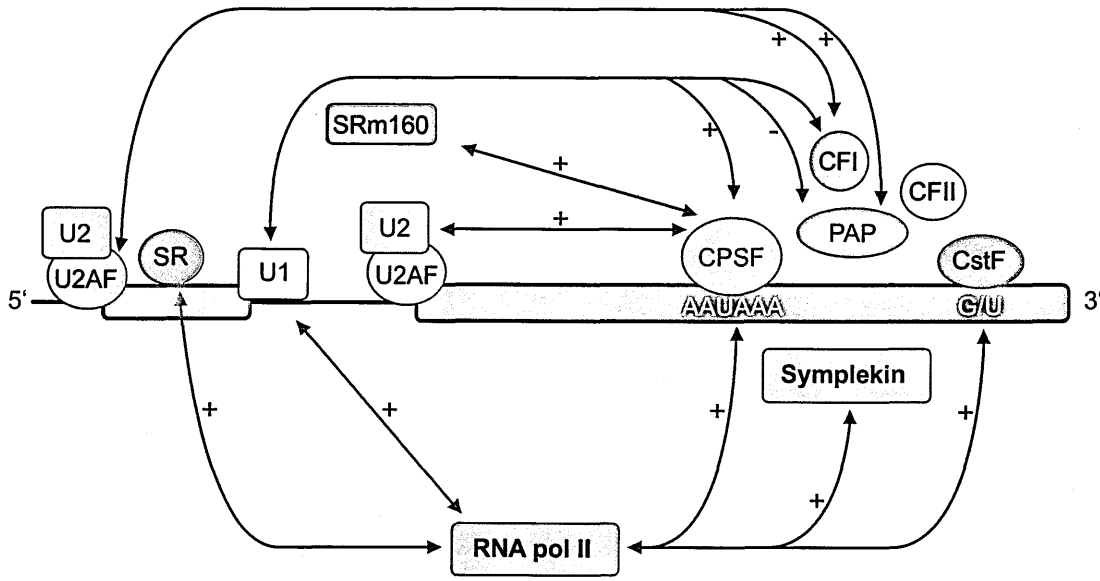


Figure 6
Crosstalk between 3'end processing and splicing factors.
For detailed explanation see text; double arrowheads indicate interactions between 3'end processing and splicing factors; plus and minus indicate positive and negative actions of factors shown, respectively; adapted from Pandya-Jones 2011.

terminal exon 3'splice site and thus acts in the stimulation of the last intron splicing (Vagner et al., 2000). In turn, another study demonstrated that binding of U2AF65 to the polypyrimidine tract of the beta globin last intron stimulates in vivo and in vitro 3'end processing (Millevoi et al., 2002). Comparably, different mutations of the polypyrimidine tract and **AG** dinucleotide at the 3'splice site of the beta globin last intron led to inefficient 3'end processing (Antoniou et al., 1998; Dye and Proudfoot, 1999). An additional *in vitro* study using as a model the adenovirus L3 polyadenylation site showed that U2AF65 exerts its 3'end processing stimulatory activity through direct interaction with the 59 kDa subunit of CFI (Millevoi et al., 2006). This interaction relies on the contact between the RS-domains present in both proteins. Similar RS-domain interactions have also been observed for the CFI-68 subunit and the SR proteins SRSF3, SRSF7 and SRSF10 (former SRp20, 9G8 and hTra2Beta, respectively) (Dettwiler et al., 2004). Another RS-domain containing protein, the SR-related protein SRm-160, is involved in the enhancement of the 3'end cleavage efficiency *in vitro* and *in vivo* (McCracken et al., 2002). This enhancement is probably due to the association of SRm-160 with CPSF. In another study, CPSF is found in close connection with the splicing factor U2snRNP and this interaction is necessary for efficient splicing and cleavage reactions in a coupled

system (Kyburz et al., 2006). Depletion of the CPSF-100 subunit from an *in vitro* splicing and cleavage coupled assay decreased not only the cleavage but also the splicing efficiency. Similarly, depletion of only U2 snRNA in the same assay led to a significant decrease in the cleavage efficiency and the complete abolishment of splicing.

The above described examples show stimulatory interactions between splicing and 3'end processing factors. However, the interactions between splicing and 3'end processing factors can be directed to the inhibition of the splicing and/or 3'end processing reactions. One of the examples of inhibition is the direct competition observed between splicing and 3'end processing reactions in the case of defining alternative terminal exons. This example of competition will be discussed below in the section Alternative processing. Several examples are known where splicing factors directly affect the cleavage and polyadenylation efficiencies. Early studies showed that U1A, a protein part of the U1snRNP complex, interacts with PAP to inhibit polyadenylation of its own transcript (Gunderson et al., 1997). Moreover, polyadenylation at the bovine papillomavirus (BPV) late polyadenylation signal is inhibited by the binding of U1snRNP to an element containing consensus 5'splice site sequence located upstream of the PAS. Mutation of this 5'splice site rescued the U1snRNP-dependent inhibition. The U1snRNP exhibited its inhibitory role on BPV polyadenylation through the interaction between its component, U1 70K, and PAP (Gunderson et al., 1998). A recent study demonstrated the role of U1snRNP in preventing premature cleavage and polyadenylation by a mechanism similar to that observed for the BPV late polyadenylation signal (Kaida et al., 2010). Depletion of U1snRNP from HeLa cells or mutation of the 5'splice site results in the activation of cryptic polyadenylation signals located downstream of the 5'splice site. These cryptic polyadenylation signal elements can be commonly found across intronic regions. Usage of such polyadenylation signals could disrupt normal processing (splicing and constitutive polyadenylation signal usage), leading to the formation of aberrant transcripts. Therefore, inhibition of 3'end processing at the cryptic intronic polyadenylation signals by U1snRNP is of high importance for the normal maturation of the pre-mRNA.

1.2 Alternative polyadenylation

As mentioned at the beginning of the Introduction section, the complexity of eukaryotic protein-coding gene expression lies mostly in the regulation of the pre-mRNA processing events. In particular, alternative processing of pre-mRNAs represents an important step in the regulation of gene expression. Alternative processing includes transcription initiation at alternative promoters, alternative splicing and alternative polyadenylation. For many years, the importance of alternative promoter and polyadenylation site usage compared to alternative splicing was underestimated. The reason for this lies in the huge potential of alternative splicing for the production of many different protein isoforms from a single gene (Black, 2003). The extreme example is the Drosophila Dscam gene, where alternative splicing of a single pre-mRNA transcript can potentially generate ~38 000 different protein isoforms. In addition, recent analysis of 15 diverse human tissues and cell line transcriptomes indicate that 92-94% of human genes undergo alternative splicing (Wang et al., 2008). However, during the last years increasing attention has been given to alternative polyadenylation. It is estimated that over half of the human and mouse genes undergo alternative cleavage and polyadenylation (Shepard et al., 2011; Tian et al., 2005). Several reviews have recently been published stressing the

growing importance of alternative polyadenylation in the regulation of gene expression (Di Giammartino et al., 2011; Lutz and Moreira, 2011; Proudfoot, 2011).

In general, there are two types of alternative polyadenylation (Figure 7). In the first type, alternative polyadenylation signals are present in different upstream introns or exons ending the message in alternative 3'terminal exons which normally results in different protein isoforms. This type represents a combination of alternative splicing and alternative polyadenylation. In the second type, multiple alternative polyadenylation signals are present in the 3'most terminal exon (tandem polyadenylation sites), leading to the formation of mRNA isoforms that differ in the lengths of their 3'UTRs. Changes in the 3'UTR length can influence mRNA stability, localization and translatability.

1.2.1 Regulatory mechanisms involved in alternative 3'terminal exon selection

The molecular mechanisms involved in the regulation of alternative polyadenylation and, thus, the recognition of specific polyadenylation sites are to date poorly understood. There are a few individual cases in which the regulatory mechanisms involved in alternative polyadenylation have been addressed in detail, such as in the well-studied immunoglobulin M (IgM) heavy chain and calcitonin/CGRP transcripts (Figure 8). Both belong to the first type of alternative polyadenylation, where alternative polyadenylation sites usage results in the presence of different terminal exons in the final transcripts. In immature B cells, IgM heavy chain pre-mRNAs are mostly polyadenylated at the distal site leading to the production of a membrane bound form of the IgM heavy chain. During differentiation to plasma cells, the secreted form of IgM (μs) is produced and the promoter-proximal cleavage-polyadenylation site is used. The membrane-associated IgM (μm) encoding mRNA is polyadenylated at the downstream "μm" PAS which is only used when the μs site is removed by alternative splicing. It has been demonstrated that the switch from the membrane bound to the secreted form is directed by the relative

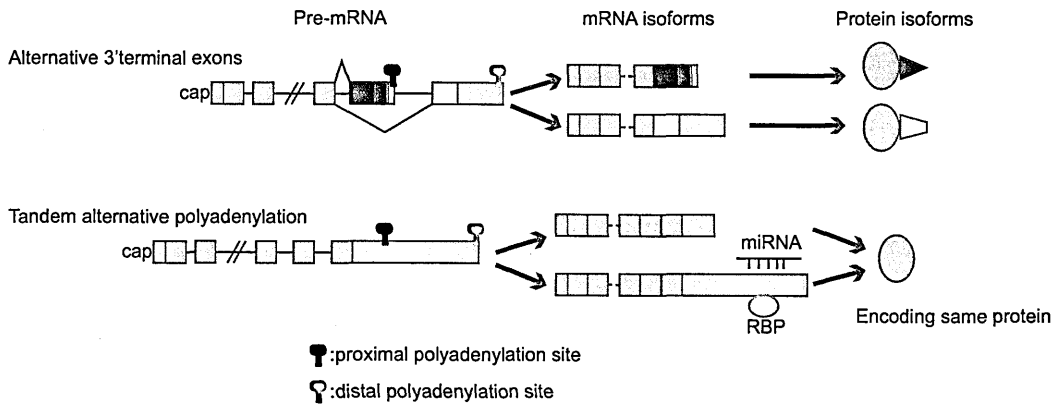
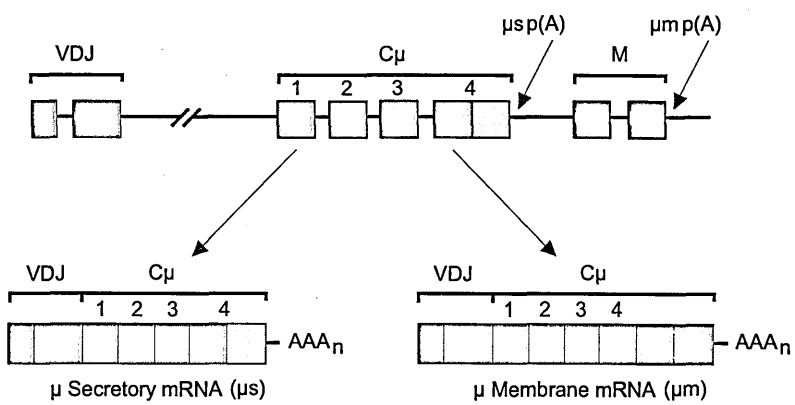


Figure 7
Types of alternative polyadenylation.
Alternative 3'terminal exon choice results in different protein isoforms; Tandem alternative polyadenylation leads to the formation of mRNA isoforms encoding same protein, but differ in the length of their 3'UTR; light blue boxes indicate common coding regions; lines represent introns; green boxes indicate noncoding regions; blue an yellow boxes indicate unshared coding regions; miRNA - microRNA; RBP - RNA binding protein; adapted from Di Giammartino et al., 2011.

IgM heavy chain



Calcitonin/CGRP

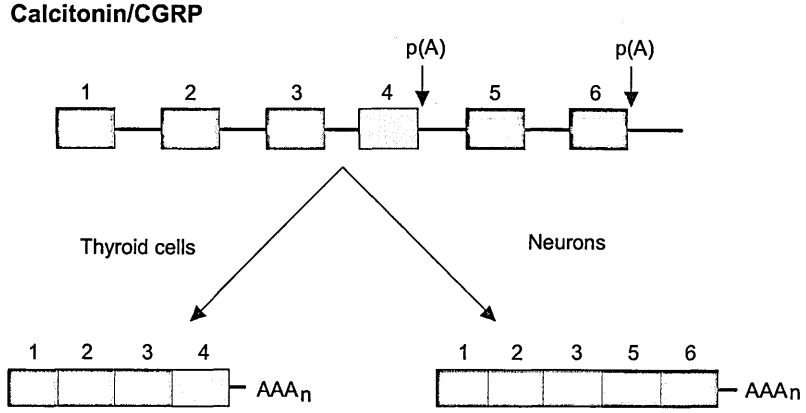


Figure 8
Alternative 3'terminal exon selection in IgM and calcitonin pre-mRNAs.
For detailed explanation see text; adapted from Zhao et al 1999.

amounts of CstF-64 and hnRNP F (Takagaki et al., 1996; Veraldi et al., 2001). CstF-64 binding affinity for the DSE of the proximal secretory-specific μs PAS is lower than for the DSE present in the distal membrane-specific polyadenylation site of the IgM heavy chain pre-mRNA. In addition, the binding site for hnRNP F is placed in the DSE of the proximal secretory-specific polyadenylation site, thus competing for access to this site with CstF-64. In B cells, high levels of hnRNP F and low levels of CstF-64 are detected, leading to the inefficient usage of the proximal secretory-specific site, its splicing and favour of the usage of the distal membrane-specific site. In plasma cells we observe the opposite situation: the levels of the hnRNP F are lower, whereas the levels of CstF-64 are higher, favouring the usage of the proximal secretory-specific poladenylation site and the expression of the secreted IgM heavy chain form. Additionaly, it has been shown that expression of the transcription elongation factor ELL2 is induced in plasma cells and that ELL2 can stimulate usage of the proximal secretory-specific poladenylation site (Martincic et al., 2009).

Another example of alternative polyadenylation through the selection of alternative 3'terminal exons is the calcitonin/CGRP pre-mRNA, where the proximal PAS is present in exon 4 and the distal PAS in exon 6, which is the 3'most terminal exon (Amara et al., 1984) (Figure 8). In neuronal tissues, the exon 4 is skipped from the calcitonin pre-mRNA and the distal polyadenylation site is used, leading to the production of the calcitonin gene-related peptide (CGRP). In thyroid C cells, the exon 4 is included and, thus, the proximal polyadenylation site is used leading to the production of the hormone calcitonin. Exon 4 inclusion and the usage of the proximal polyadenylation site of the calcitonin pre-mRNA are stimulated by the presence of a 127 nt long enhancer element. The enhancer element is located in the intronic region, 150 nt downstream of exon 4. The splicing factors PTB, SRSF3 (former name SRp20) and U1snRNP bind the enhancer element and positively regulate the inclusion of exon 4 and the usage of the proximal polydenylation site of the calcitonin/CGRP pre-mRNA (Lou et al., 1999; Lou et al., 1998).

Both, the IgM heavy chain and calcitonin/CGRP pre-mRNAs contain alternative exonic

polyadenylation signals. Examples of intronic alternative polyadenylation signals are found to be common in the gene expression regulation of the receptor tyrosine kinase (RTK) (Vorlova et al., 2011). Usage of the intronic alternative polyadenylation signals in the RTK pre-mRNAs leads to the generation of soluble truncated RTK isoforms, which lack the C-terminal trans-membrane domain. The soluble RTK isoforms preserve the binding properties for the ligand and are important for the dominant-negative regulation of the signaling pathway, and are termed soluble decoy RTK isoforms (sdRTK). The activation of the intronic alternative polyadenylation signals of RTK pre-mRNAs can be achieved through the inhibition of U1snRNP binding to the upstream 5'splice site (Vorlova et al., 2011). This result is consistent with the above-mentioned role of U1snRNP in the inhibition of the premature cleavage and polyadenylation in the introns (Kaida et al., 2010) and highlights the importance of U1snRNP in the regulation of intronic alternative polyadenylation. Vorlova and collaborators demonstrated that activation of the sdRTK isoforms by blocking U1snRNP is important not only from the mechanistic point of view but also could be a potent therapeutic strategy for a number of diseases which involve aberrant RTK signaling.

1.2.2 Regulatory mechanisms involved in tandem alternative polyadenylation site selection

Although the number of individual examples of tandem alternative polyadenylation is constantly increasing, the molecular mechanisms that regulate these events remain largely uncharacterized. As described below, there are a few isolated studies proposing that different trans-acting factors can regulate the usage of tandem alternative polyadenylation sites.
In addition to the above described alternative polyadenylation regulation in the IgM heavy chain pre-mRNA, CstF-64 levels can also regulate the usage of other alternative polyadenylation sites present in the 3'UTR of genes. Overexpression of CstF-64 in LPS-stimulated macrophages leads to a shift in the usage of the distal to the proximal polyadenylation sites in a number of genes (Shell et al., 2005).

In similar fashion, knock-down of CFI-25 in HeLa cells influences the switching in the usage of tandem alternative polyadenylation sites present in the 3'UTR of TIMP-2, syndecan2, ERCC6 and DHFR transcripts (Kubo et al., 2006). For all the analyzed genes the distal site(s) were preferentially used in HeLa cells, and CFI-25 depletion caused a shift from the usage of the distal to the proximal polyadenylation sites in these genes.

Very recently, the nuclear factor PABP1 (PABPN1), another component of the 3'end processing machinery, was shown to be important for polyadenylation site selection. Using a deep-sequencing-based approach tailored to the analysis of 3'end formation, the effects of PABPN1 knock-downs on polyadenylation site selection were analyzed. The knock-downs were performed in human cancer (U2OS - osteosarcoma) and non-transformed (RPE-1 - retinal pigment epithelial) cells. The authors identified 572 and 357 transcripts in U2OS and RPE-1 cells, respectively, showing a shift in polyadenylation site usage in the absence of PABPN1 (Jenal et al., 2012). More than 90% of the detected transcripts in U2OS cells and 82% in RPE-1 cells showed a shift towards proximal polyadenylation sites, a trend similar to that observed in CFI-25 depleted cells (Kubo et al., 2006). Furthermore, the authors proposed that PABPN1 directly inhibits 3'end processing at proximal polyadenylation signals. Short expansions of GCG triplet repeats in the PABPN1 gene (trePABPN1) are the cause of oculopharyngeal muscular dystrophy (OPMD). 3'Seq analysis of skeletal muscle from a mouse model of OPMD, expressing the trePABPN1 allele, demonstrated the same shift in polyadenylation site selection as was detected in cell lines artificially depleted of PABPN1.

NOVA2, a neuronal-specific splicing factor known to influence alternative splicing, was shown to be also implicated in the regulation of alternative polyadenylation. Affymetrix exon arrays of mRNAs derived from NOVA2 wild-type or knockout mouse brains led to the identification of 297 transcripts that showed differences in their 3'UTR length (Licatalosi et al., 2008). NOVA2 binding sites were found to be enriched around polyadenylation sites. The effect of NOVA2 on polyadenylation site usage was shown to be positional dependent. Stimulation of polyadenylation site usage was observed when the binding

sites of NOVA2 were positioned upstream of the polyadenylation sites. In contrast, an inhibitory role was observed when NOVA2 binding sites were present downstream of the polyadenylation sites. For example, in the case of the Cugbp2 and Slc8a1 genes it was shown that NOVA2 binding sites are enriched downstream of the proximal polyadenylation sites of their transcripts. NOVA2 knockout resulted in a shift from the distal towards proximal polyadenylation site usage in both genes.

The **AU**-rich element (ARE) binding protein HuR is expressed ubiquitously and is critical in the regulation of mRNA stability. Recently, it has been demonstrated that HuR autoregulates its expression by influencing alternative polyadenylation site usage in its own transcript (Dai et al., 2012). Elevated levels of HuR in the nucleus stimulate its binding to the **GU**-rich DSE of the proximal polyadenylation site in the HuR pre-mRNA. Thus, HuR is able to compete with the binding of CstF-64 to the proximal polyadenylation site favouring usage of the distal site. Thus, high levels of HuR result in the formation of the long HuR mRNA isoform. In contrast to the short, this long isoform contains destabilizing AU-rich elements in its 3'UTR, increasing mRNA decay rates and so result in a reduction of HuR protein production. Moreover, it has also been found that three neuron-specific Hu family members influence alternative polyadenylation of the HuR pre-mRNA in the same manner as HuR, leading to the production of the long HuR mRNA isoform and the decrease HuR levels in neurons (Mansfield and Keene, 2012). Interestingly, the ELAV protein, a Drosophila homolog of the mammalian Hu family, was shown to regulate alternative 3'terminal exon selection of the ewg pre-mRNA in neurons (Soller and White, 2003). The usage of the ewg pre-mRNA intronic polyadenylation site is inhibited by ELAV binding to the AU-rich elements downstream of that site, leading to intron splicing and polyadenylation at the 3'most terminal exon. However, contrary to the HuR protein, ELAV does not affect dCstF-64 binding to the DSE of the intronic polyadenylation site present in the ewg pre-mRNA (Soller and White, 2003).

Interestingly, another example of polyadenylation in Drosophila revealed that RNAP II elongation kinetics can influence site usage of tandem alternative polyadenylation sites

(Pinto et al., 2011). The polyadenylation pattern in Drosophila was analyzed in strains containing a mutant version of the RNAPII which shows 50% decrease in the transcription elongation rates. A shift from distal to proximal polyadenylation site usage has been observed in mutant flies for six out of the eight analyzed genes.

1.2.2.1 3'Untranslated region (3'UTR)

As already mentioned, the usage of different tandem polyadenylation sites leads to the formation of mRNAs with the same coding capacity but different 3'UTR length. A variety of cis-regulatory elements can be found along the 3'UTR which can influence the mRNA stability, cytoplasmic localization and translational efficiency (Figure 9). Usage of the tandem alternative polyadenylation sites frequently depends on the tissue type, the developmental stage, state of the cell growth and differentiation, and also different pathological conditions, like cancer. For example, during proliferation, dedifferentiation or cell transformation, the usage of proximal polyadenylation sites is stimulated leading to the formation of mRNAs with shorter 3'UTRs (Ji and Tian, 2009; Mayr and Bartel, 2009; Sandberg et al., 2008). Shorter 3'UTRs contain less regulatory elements, which are associated with increased protein production (Mayr and Bartel, 2009; Sandberg et al., 2008). On the contrary, during differentiation or development, the usage of distal polyadenylation sites is favored leading to the formation of mRNA isoforms containing longer 3'UTRs (Ji et al., 2009). Longer 3'UTRs contain more regulatory elements that are able to cause destabilization of mRNAs and reduction in their translational efficiency, leading to a decrease in protein production (Mayr and Bartel, 2009; Sandberg et al., 2008). Regulatory elements located along the 3'UTR are recognized by different RNA-binding proteins and/or microRNAs. For example, the previously mentioned members of the mammalian Hu family are **AU**-rich element (ARE) binding proteins which stimulate mRNA stability. Moreover, there are other ARE binding proteins that can provoke mRNA destabilization and degradation, such as tristetraproline (TTP) (von Roretz et al., 2011). **AU**-rich elements are usually represented by one or several **AUUUA** pentamers

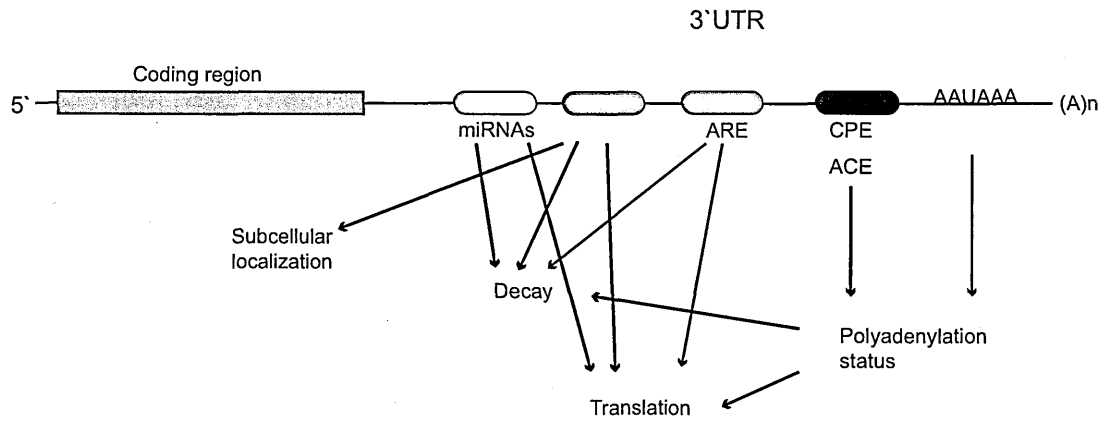


Figure 9
Schematic representation of the cis-acting regulatory elements present in the 3'UTR.
Red box indicates microRNAs (miRNAs) target sites; green box indicates cis-acting elements that are recognized by the trans-acting factors other then CPEB and ARE-binding proteins; blue box indicates AU-rich element (ARE); violet box indicates cytoplasmic polyadenylation element (CPE) and adenylation control element (ACE); AAUAAA indicates the position of polyadenylation signal; arrowheads indicate various pathways by which cis-acting regulatory elements present in the 3'UTR influence gene expression; adapted from Lutz and Moreira 2011.

or **UUAUUUA(U/A)(U/A)** nonamers. Another element that can be present in the 3'UTR is the cytoplasmic polydenylation element (CPE). The consensus sequence of CPE is the **UUUUAU** motif, which is bound by the cytoplasmic polyadenylation element binding protein (CPEB) (Mendez and Richter, 2001). Association of CPEB with the CPE leads to translational repression of the mRNAs. Phosphorylation of the CPEB results in its dissociation from the CPE, triggering cytoplasmic polyadenylation and translational activation of the mRNA. The impact of the CPEB translational repression depends on the position and number of CPEs in respect to the hexanucleotide motif and pumilio-binding element (PBE) (Pique et al., 2008). The PBE is another important element implicated in translational repression. The **UGUANAUA** sequence is the PBE consensus motif and represents one of the 50 most conserved octamers present in the mammalian 3'UTRs (Xie et al., 2005). Members of the PUF protein family (**Pu**milio protein in *Drosophila melanogaster* and ***fem***-3 binding factor protein in *Caenorhabditis elegans*) recognize the PBEs and stimulate translational repression. There are two members of mammalian PUF proteins, called PUM1 and PUM2. The PUM2 is expressed in hippocampal neurons localizing to dendrites and its depletion leads to altered dendritic outgrowth and synaptic function (Vessey et al., 2010). The CPEB was also shown to regulate translation of mRNAs localized in dendrites of mammalian neurons, such as CAMKIIα mRNA. In addition to its role in translational repression, CPEB was shown to affect the dendritic targeting of the CAMKIIα mRNA (Huang et al., 2003). Other proteins are also found to sti-

mulate targeting of mRNAs to dendrites, such as ZBP1 and Staufen. These proteins bind localization elements placed along the 3'UTR of the target mRNAs.

As mentioned above, some of the 3'UTR sequence motifs are targets for microRNAs (miRNAs). miRNAs are ~22 nucleotides long noncoding RNAs. 6-8 nucleotides of the 5'end of the miRNAs represent the seed region which binds imperfectly complementary elements in the 3'UTR of their target mRNA (Bartel, 2004). Binding of the miRNAs to their target sequences results in mRNA degradation and repression of translation. Translational repression by miRNAs is a potent and widespread mechanism of gene expression regulation. According to the latest realase of the constantly growing miRNA database, miRBase (Release 19, August 2012, http://www.mirbase.org), there are 1600 precursors and 2042 mature human miRNAs. It has been estimated that miRNAs regulate the expression of more than half of the mammalian genes (Friedman et al., 2008). Therefore, miRNAs are, directly or indirectly, implicated in all cellular pathways, including RNA processing. For example, splicing outcome can be influenced by the biogenesis of intronic miRNAs (Janas et al., 2011; Shomron and Levy, 2009), or by direct miRNA targeting of splicing factors, such as PTB (Makeyev et al., 2007). The neuronal specific microRNA mir-124 directly targets PTB, a repressor of neuronal specific splicing, causing the increase in the inclusion of neuron-specific alternative exons and enhancing neuronal diferentiation (Makeyev et al., 2007). The miRNAs exert their role in the cytoplasm. However, certain miRNAs can be localized in the nucleus (Hwang et al., 2007; Liao et al., 2010; Jeffries et al., 2011), where they can mediate transcriptional gene silencing (TGS) (Malumbres, 2012). For example, mir-320 was shown to mediate TGS at the POLR3D promoter (Kim et al., 2008). It was also shown that nuclear miRNAs can target antisense non-coding RNAs, as in the case of nuclear mir-671, which directly downregulates circular antisense transcript of the CDR1 gene (Hansen et al., 2011). The antisense CDR1 non-coding RNA positively regulates CDR1 mRNA levels and its downregulation leads to a decrease in the CDR1 mRNA levels. As miRNAs target 3'UTRs, it would be interesting to study the possible role of the nuclear miRNAs in the regulation of 3'end processing.

1.3 Beta adducin

Beta adducin is a member of the adducin protein family. In mammals adducin protein family comprises of alpha, beta and gamma adducins which are encoded by the ADD1, ADD2 and ADD3 genes, respectively. Adducin is a membrane skeleton protein and functions as a dimer-tetramer composed of alpha/beta or alpha/gamma heterodimers. It is localized at the spectrin-actin junctions where it has a role in capping the fast-growing ends of actin filaments, bundling actin filaments and recruiting spectrin to actin (Matsuoka et al., 2000).

The expression of beta adducin is restricted to brain and hematopoietic tissues, whereas alpha and gamma subunits are expressed ubiquitously (Gilligan et al., 1999). Several alpha and beta adducin polymorphisms have been found to be associated with hypertension in humans and Milan hypersensitive rats (Cusi et al., 1997; Tripodi et al., 1991). In addition, beta-adducin-deficient mice suffer from mild anemia with compensated hemolysis, which corresponds to the human spherocytic hereditary elliptocytosis syndrome (Muro et al., 2000). Moreover, absence of beta adducin in the knockout (KO) mice causes motor coordination, behavioral and learning deficits which are associated with impairment of long-term potentiation (LTP) and complete abolishment of long-term depression (LTD) (Porro et al., 2010). Beta adducin is involved in the stabilization of synaptic structures in mice housed in an enriched environment, containing different types of stimuli such as toys and running wheels, which lead to enhanced plasticity (Bednarek

and Caroni, 2011). Upon environmental enrichment, increased levels of phosphorylated beta adducin in mice hippocampus influence the disassembly of synapses, whereas non-phosphorylated beta adducin increases the assembly of new synapses (Bednarek and Caroni, 2011). Therefore, upon enhanced plasticity the absence of beta adducin in KO mice leads to increased synapse turnover and impairment of long-term memory and learning (Bednarek and Caroni, 2011).

1.3.1 Mouse beta adducin gene structure and expression

The mouse beta adducin gene is located on chromosome 6, spans about 100 Kbp and comprises 16 exons (Figure 10). Tissue-specific expression of the mouse beta adducin gene is predetermined by the use of alternative tissue-specific promoters and first exons (Costessi et al., 2006). The gene promoter and the first exon that are specific for neuronal tissues (brain promoter and first exon) are located roughly 50 Kbp upstream of the ones specific for erythropoietic tissues (spleen promoter and first exon). Tissue specificity of the brain promoter is only observed in mice and rats, whereas in human its activity is also observed in spleen and bone marrow, as revealed by the RT-PCR analyses of the first exon expression in different human, mouse and rat tissues (Costessi et al., 2012). Activation of the spleen promoter is detected only in erythropoietic tissues for all three species. The translation start site is located in the second common exon, thus the presence of the tissue-specific first exons is not involved in the structure of the final protein product. In addition to the tissue specificity of the promoter usage, tissue specific mRNA isoforms are created by the usage of three tandem alternative polyadenylation sites placed in the 3'most terminal exon of the mouse beta adducin pre-mRNA (Costessi et al., 2006). In mice, the use of the proximal polyadenylation sites (PAS1 and PAS2-3) in erythroid tissues generates 3'UTRs in length of ~800 and ~1300 bases, respectively, while the use of the distal polyadenylation site (PAS4) in brain generates an unusually

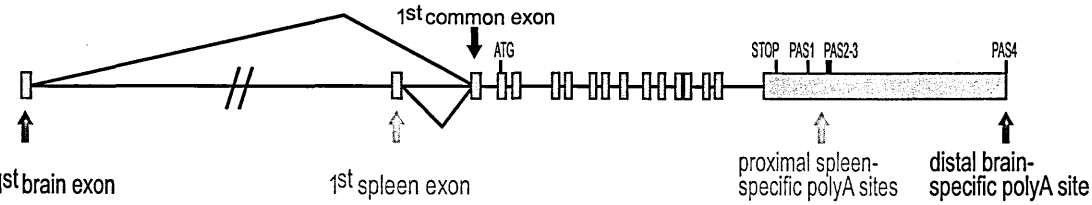


Figure 10
Structure of the mouse beta adducin gene.
Adapted from Costessi et al. 2006

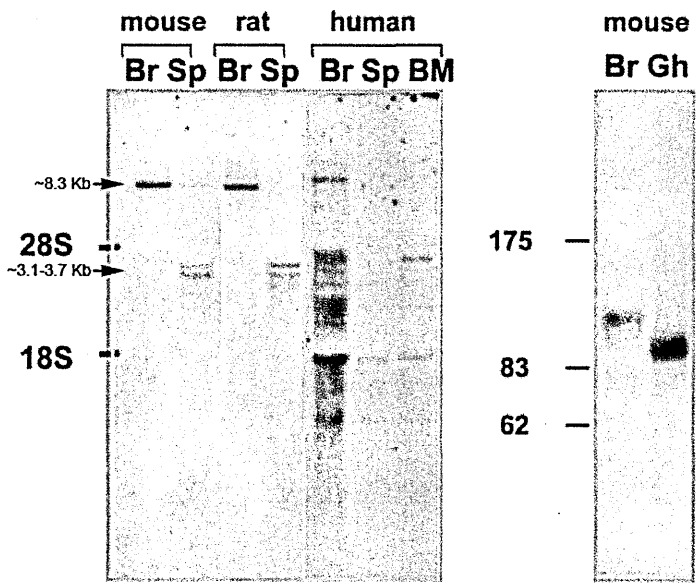


Figure 11
Tissue specific expression of the short and long beta adducin mRNA isoforms.
Left panel – Northern blot analysis of RNAs from mouse, rat and human brain and spleen, and human bone marrow (BM); Right panel – Western blot analysis of mouse brain and ghost protein extracts. Taken from Costessi et al. 2006.

long 3'UTR of about 5.7 kilobases. The presence and use of the different tandem alternative polyadenylation sites of the beta adducin is very well conserved among different species suggesting an important and tissue specific roles for each of the generated 3'UTRs possibly by affecting mRNA stability, translatability and subcellular localization. To summarize, in mouse erythroid tissues, beta adducin gene expression is determined by the usage of the spleen specific promoter and the proximal polyadenylation sites leading to the formation of a ~3.1-3.7 Kb long mRNA isoform (Figure 11, left panel, (Costessi et al., 2006)). In contrast, in mouse brain, an ~8.3 Kb long beta adducin mRNA isoform is formed, due to the usage of the brain specific promoter and the distal

polyadenylation site. Both mRNA isoforms lead to the production of a protein that is similar in size in mouse erythroid tissues (97 kDa) and brain (110 kDa) (Figure 11, right panel, (Costessi et al., 2006)). The molecular mechanism involved in the regulation of the beta adducin tissue specific alternative promoter and polyadenylation site usage to date are not characterized.

To complicate matters, beta adducin pre-mRNA is also subjected to alternative splicing and alternative 3'terminal exon selection. Alternative splicing of several internal exons of the beta adducin pre-mRNAs leads to the formation of different mRNA variants. However, these transcript variants represent a low fraction of the total beta adducin mRNA pool and the corresponding proteins have not been detected. In addition, intron 13 retention and usage of the polyadenylation site placed in this intron leads to the formation of shorter mRNA transcripts (beta-add63 mRNAs). Although low levels of the beta-add63 mRNAs can be detected by Northern blots using RNAs isolated from mouse spleen and brain (Costessi et al., 2006), a corresponding protein has not yet been detected.

1.3.1.1 Mouse beta adducin tandem alternative polyadenylation signals (structure and conservation)

As previously mentioned, there are three polyadenylation signals placed in the mouse beta adducin 3'UTR. The conservation between humans, mice and rats and the predicted cis-acting elements of each signal are described in Costessi et al., 2006. Mice and rats contain two proximal polyadenylation signals, whereas three sites are present in humans. All species contain only one distal polyadenylation signal.

The region of the first proximal polyadenylation signal shows modest sequence conservation between humans, mice and rats (Figure 12). The cleavage and polyadenylation site (PAS1) is represented by a **G**uanosine at the base position 1076, relative to the first base of the mouse beta adducin last exon. A canonical **AUUAAA** hexanucleotide motif is found 18 nucleotides upstream of the cleavage site. The observed hexanucleotide motif is poorly conserved, as in the humans the noncanonical **ACUAAA** motif is present. Also the

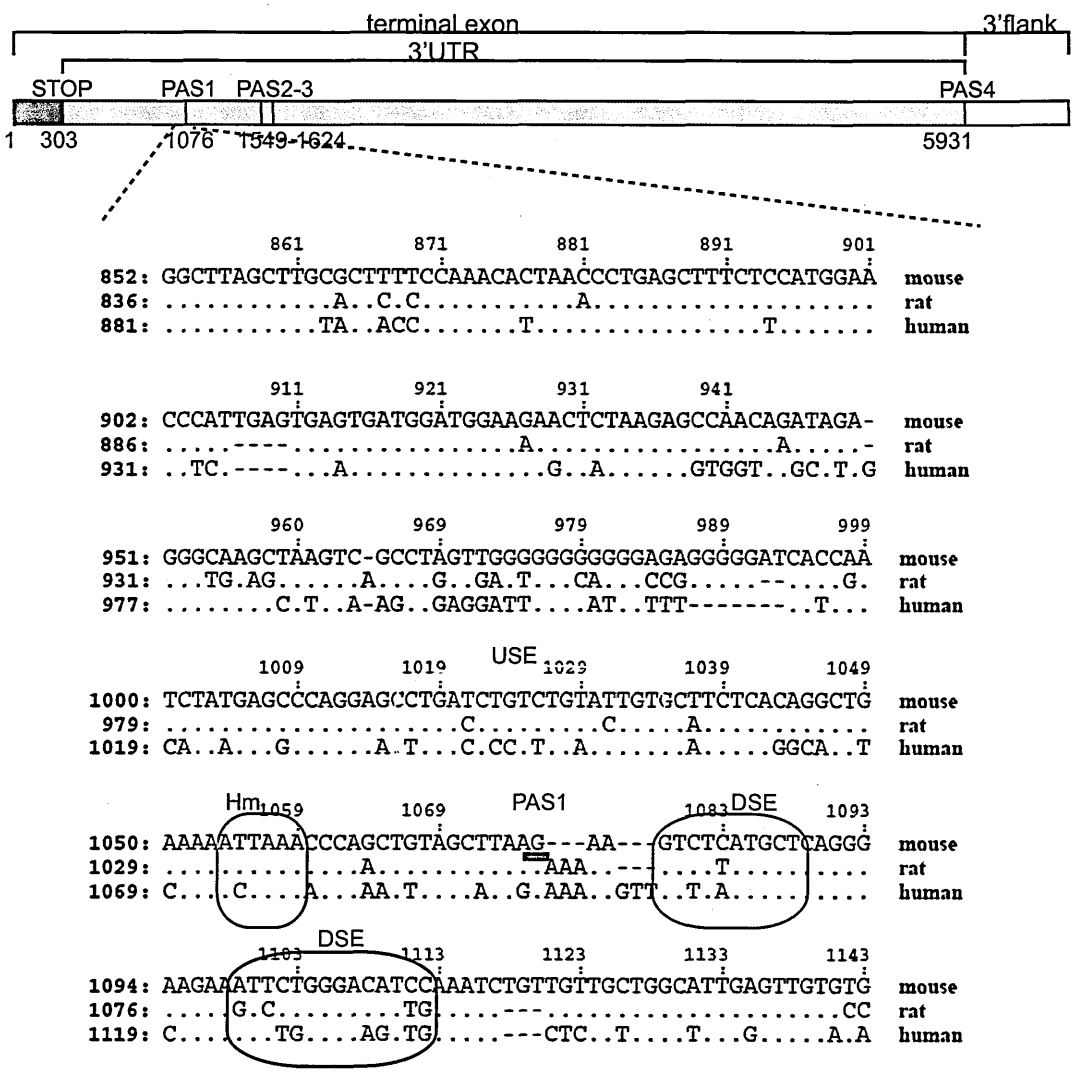


Figure 12
Sequence structure and conservation of the region containing mouse beta adducin first proximal polyadenylation signal.
Upper panel – schematic representation of the mouse beta adducin last exon; numbers indicate nucleotide positions; a dark grey box indicates the coding region; a light gray box indicates the 3'UTR; a white box indicates 3'genomic flanking region; STOP – stop codon; PAS1 – the first polyadenylation site; PAS23 – the second polyadenylation site; PAS4 – the fourth polyadenylation site; the alignment of the genomic sequences of humans, rats and mice in the first proximal polyadenylation region is enlarged below; MultiPipMaker global sequence alignment program was used; USE – upstream sequence element; Hm – hexanucleotide motif; DSE – downstream sequence element.

putative **U/GU**- rich upstream and downstream sequence elements are poorly conserved. The second proximal polyadenylation signal contains multiple cleavage sites that are located in the region from the base 1549 to the base 1624, relative to the first base of the mouse last exon (Figure 13). Cleavage at either of these sites is considered as the usage of the second proximal polyadenylation site (PAS2-3). None of the described hexanucleotide motifs (Tian et al 2005) are present in the PAS2-3 region. The alignment of the second proximal polyadenylation signal has shown poor conservation among the analyzed species.

Contrary to the proximal polydenylation signals, the distal polyadenylation signal is well conserved between humans, mice and rats (Figure 14). Moreover, comparison of the beta adducin last exon genomic sequences in 30 vertebrate species showed two major islands of conservation (Figure 15). The first island represents the coding region of the last exon, which was expected as the coding regions are highly conserved. The second island of conservation is located in the genomic region containing the distal polyadenylation signal. A high degree of conservation has been found in all the analyzed polyadenylation elements: the hexanucleotide motif, the cleavage site and the **U/GU** rich upstream and downstream sequence elements (Figure 14). The cleavage and polyadenylation site of the distal polyadenylation signal (PAS4) is represented by the **CA** dinucleotide, the most common dinucleotide found for vertebrate cleavage sites (Sheets et al., 1990). The **CA** dinucleotide is positioned at the base 5931, relative to the starting base of the mouse beta adducin last exon. The putative hexanucleotide motif is located 20 bases upstream of the cleavage site and is represented by the **AGUAAA** variant, which is found as the polyadenylation hexamer in ~4% of the human and mouse genes (Tian et al., 2005). The absence of the canonical and more efficient **AAUAAA** hexanucleotide motif is probably compensated by the presence of the highly conserved **U/GU** rich upstream and downstream elements. Additionally, about 200 bases upstream of the cleavage site, in the highly conserved and **AU**-rich region, one putative ARE

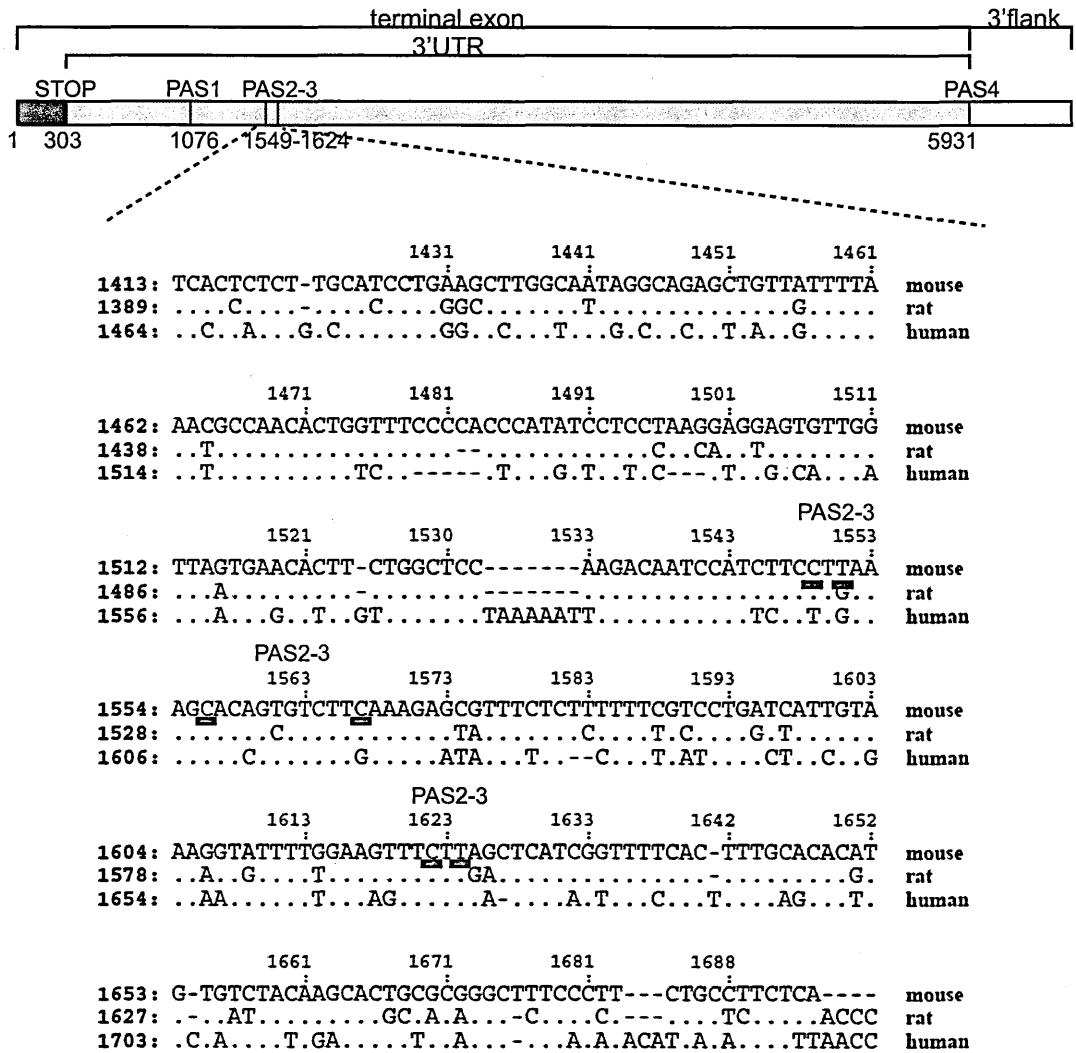


Figure 13
Sequence structure and conservation of the region containing mouse beta adducin second proximal polyadenylation signal.
Upper panel – schematic representation of the mouse beta adducin last exon; indications are as described in Figure 12; the alignment of the genomic sequences of humans, rats and mice in the second proximal polyadenylation region is enlarged below; MultiPipMaker global sequence alignment program was used.

and two putative CPE motifs are found.

Overall, the distal polyadenylation signal of the mouse beta adducin and its flanking regions show a higher level of conservation, suggesting that this signal could be more efficiently used than the proximal ones. In general, distal signals are stronger than the proximal ones in an array of tandem polyadenylation signals (Beaudoing et al., 2000; Legendre and Gantheret, 2003). Diverse sequence composition and conservation rate of the mouse beta adducin polyadenylation signals could acount for the alternative polyadenylation regulation.

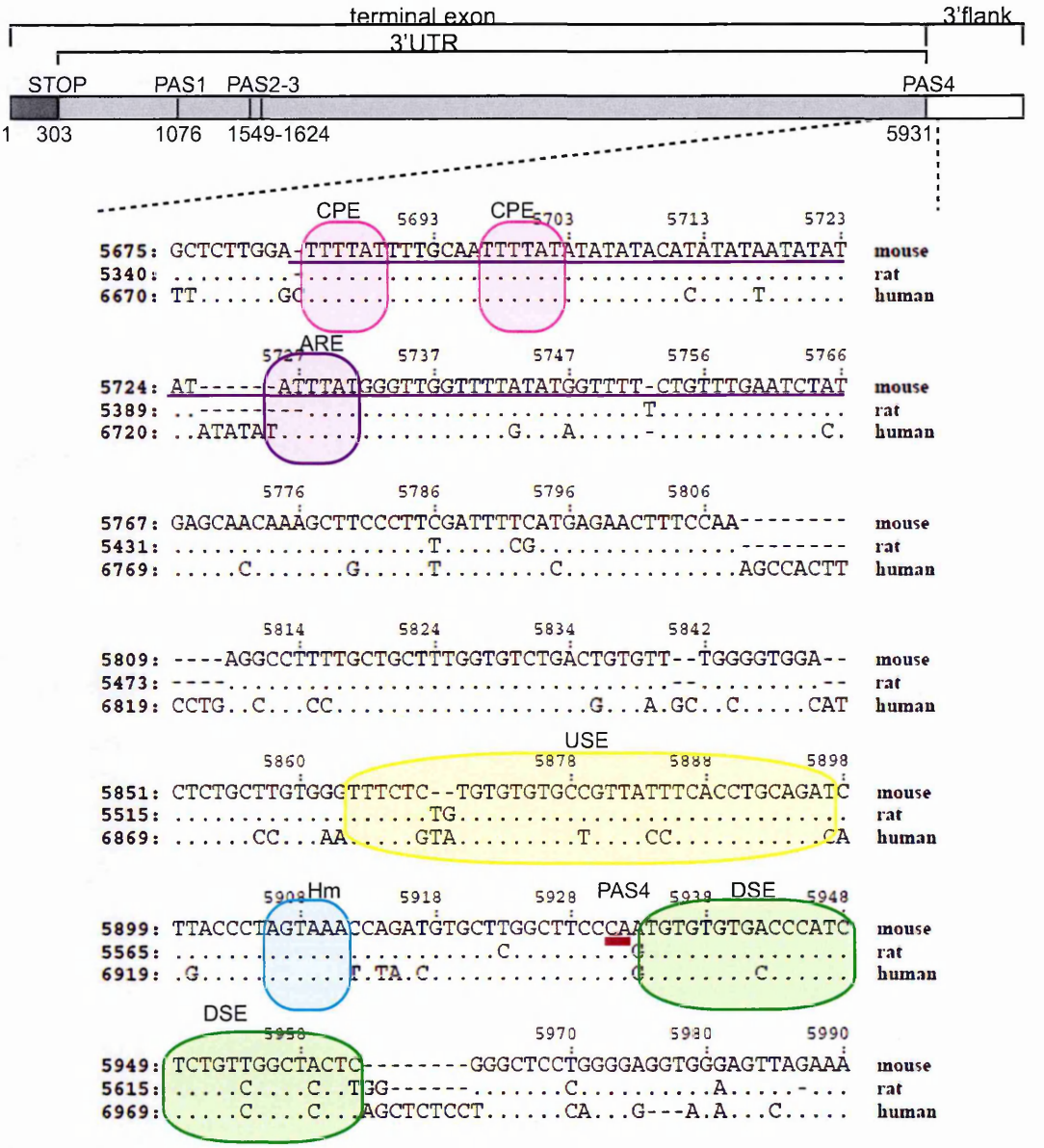


Figure 14

Sequence structure and conservation of the region containing mouse beta adducin distal polyadenylation signal. Upper panel – schematic representation of the mouse beta adducin last exon; indications are as described in Figure 12; the alignment of the genomic sequences of humans, rats and mice in the distal polyadenylation region is enlarged below; MultiPipMaker global sequence alignment program was used; CPE – cytoplasmic polyadenylation element; ARE – AU-rich element.

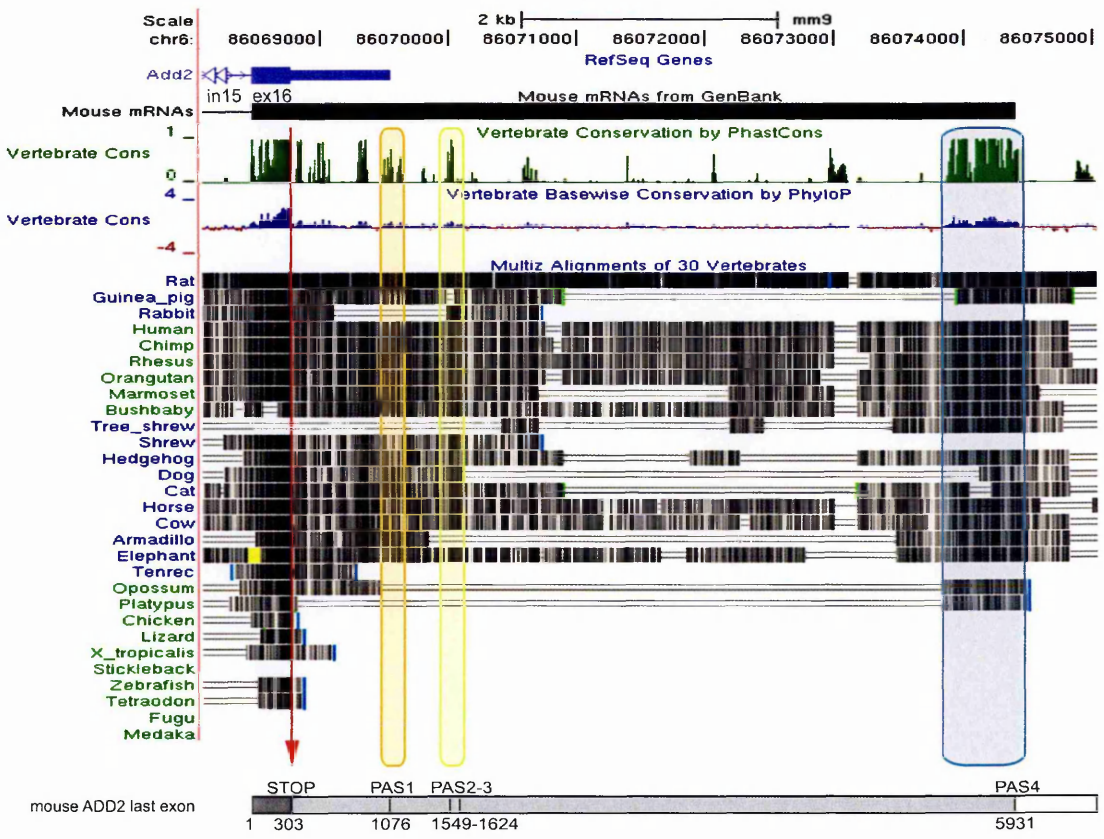


Figure 15
Conservation along the mouse beta adducin last exon between 30 vertebrate species
The data were obtained from the UCSC genome browser; lower panel – schematic representation of the mouse beta adducin last exon, indications are described in Figure 12

Aims of the thesis

As already mentioned, the beta adducin pre-mRNA undergoes tissue specific alternative polyadenylation. Two proximal spleen-specific and one distal brain-specific polyadenylation sites are placed along the 3'UTR of the mouse beta adducin gene. In the mouse brain, the usage of the distal polyadenylation site leads to the formation of an mRNA isoform with an unusually long 3'UTR (~5.7 Kb), which can contain different cis-regulatory elements regulating stability, localization and translability of the beta adducin mRNA.

In order to understand the molecular mechanisms regulating alternative polyadenylation of the mouse beta adducin pre-mRNA, the main aim of the thesis is the identification of cis-acting elements, and possible trans-acting factors, that regulate the usage of the distal polyadenylation site (PAS4) in the pre-mRNA. In addition, we aim to clarify whether the long 3'UTR harbors regulatory sequences that influence mRNA stability and translational efficiency.

Chapter 2

Results

2.1 Identification of the canonical cis-acting elements regulating usage of the distal PAS4 of the beta-adducin pre-mRNA

The beta-adducin pre-mRNA can be cleaved and polyadenylated at different PASs in a tissue-specific manner. However, it is certainly unclear if alternative polyadenylation plays a physiological role in the regulation of this gene. We therefore, aim to establish to what extent alternative polyadenylation affects beta-adducin gene expression. Understanding the regulatory mechanisms involved in this alternative pre-mRNA processing is the first step to address this question. To do so, it is important to identify the cis-acting sequences and characterize the trans-acting factors that direct alternative polyadenylation in the beta-adducin pre-mRNA. An effective strategy to study this issue is based in the generation of chimeric reporter minigenes in which the genomic regions of interest are cloned downstream of an alpha globin based reporter construct (See Materials and Methods Section). We therefore cloned the three PASs present in the mouse beta adducin 3'UTR and their respective flanking sequences downstream of the alpha globin gene, effectively substituting the alpha globin PAS. The final construct and the selected beta adducin sequences included are represented in Figure 16. The first region, called **A1**, contained the PAS1 spanning from base 17 to 1416, respective to the first base of the beta adducin last exon. The region called **A23** included the PAS2-3

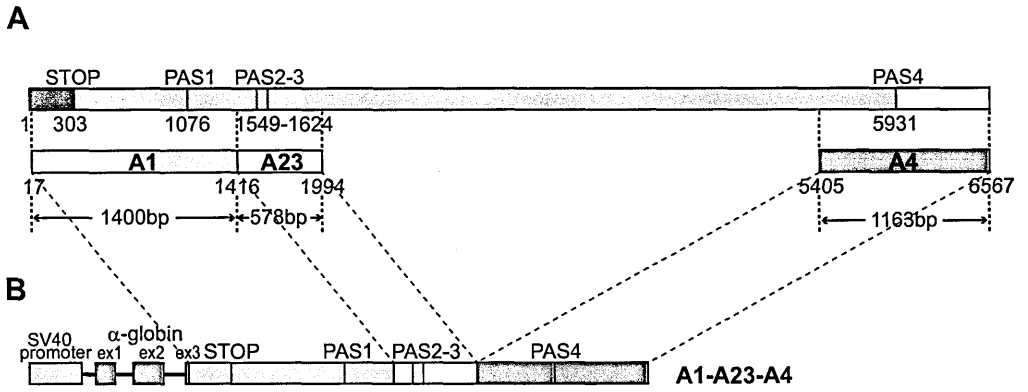


Figure 16
Schematic representation of the generation of the mouse beta adducin chimeric minigene construct.
A: schematic representation of the mouse beta adducin last exon as shown in Figure 12.; different regions (**A1**, **A23** and **A4**) containing polyadenylation sites mouse beta adducin polyadenylation sites (PAS1, PAS23 and PAS4) are represented as orange, yellow and blue boxes, respectively; arrowheads indicate the size of each region; **B**: schematic representation of the mouse beta adducin chimeric minigene construct containing all three polyadenylation sites, **A1-A23-A4**; green box indicates SV40 promoter region; violet boxes indicate exons of the alpha globin gene; STOP indicates the position of the beta adducin stop codon, which is also the expected stop codon of the globin-adducin chimera.

and extended from base 1416 to 1994. The last region, called **A4**, contained the PAS4 and spanned from base 5405 to 6567. The final construct, called **A1-A23-A4**, contained all three regions (**A1**, **A23** and **A4**) (Figure 16B). The chimeric gene was under the transcriptional control of the SV40 promoter. Due to its high transfection efficiency rate, the HeLa cell line was the first one in which expression of the **A1-A23-A4** construct was analyzed. As it can be seen in Figure 17B, only one signal was obtained in the Northern blot analysis. The size of the detected band suggested that the PAS used was the PAS4 (the expected size was 2.93Kb without the polyA tail). To confirm this hypothesis, a 3'RACE PCR experiment was performed, followed by sequencing of the amplified PCR products (Figure 17A and 17C). This experiment demonstrated that pre-mRNA processing in HeLa cells of the transcript originating from the **A1-A23-A4** construct reproduced the use of the PAS4, as it normally occurs for the endogenous beta-adducin gene in neuronal tissues (Costessi et al., 2006). In mouse and rat brain, the majority of the beta adducin transcripts is processed at the PAS4, whereas a small fraction of transcripts ends at proximal PASs (PAS1 and PAS2-3) (Costessi et al., 2006; Tripodi et al., 1991). The less efficient 3'end procesing of the PAS1 and PAS2-3 in the brain could be due to their supression by some factors and/or a lack of specific factors that are present in the erythroid tissues, where these PASs are efficiently used. Consequently, the supressing

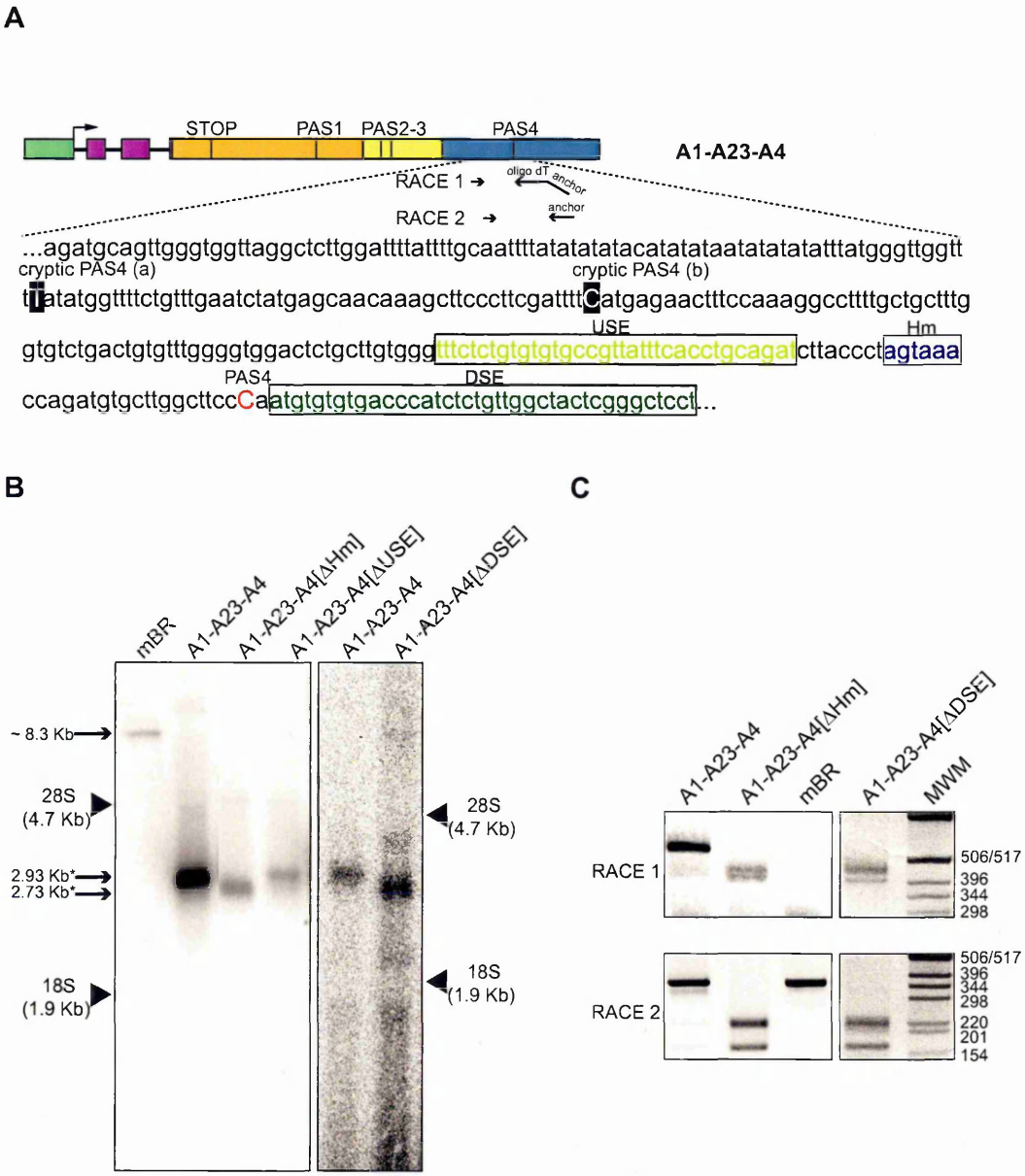


Figure 17

Identification of the core elements of the mouse beta adducin distal polyadenylation signal.

A: schematic representation of the mouse beta adducin chimeric minigene construct, **A1-A23-A4**; indications are as described in Figure 16; region containing distal polyadenylation signal is enlarged below; arrowheads indicate the primer pairs (I and II) used for the 3'RACE analysis, panel C; **B**: Northern blot analysis of 15 µg total RNA prepared from mouse brain (mBR) and 1,5 µg total RNA prepared from HeLa cells previously transfected with the **A1-A23-A4**, **A1-A23-A4[**ΔHm**]**, **A1-A23-A4[**ΔUSE**]** and **A1-A23-A4[**ΔDSE**]** constructs; the positions of the 28S and 18S rRNAs are indicated with black arrows; the sizes of the mRNAs are indicated on the left; there is no control for the transfection efficiency. **C**: 3'RACE analysis of the RNAs used for the Northern blot showed in panel B; the first (I) and the second (II) 3'RACE PCR reactions are indicated on the left; MWM – molecular weight marker.

factors could direct the polyadenylation machinery to the distal PAS4. However, the mechanisms regulating 3'end processing of the beta adducin transcripts are not known. Considering the fact that polydenylation of the beta-adducin pre-mRNA occurs in a tissue-specific manner, expression of the **A1-A23-A4** construct has been tested in different cell lines of neuronal (SK-N-BE, SH-SY5Y and PC12 cells) or erythroid (MEL and K562 cells) origin. Unfortunately, no reporter signal was detected in these cells, probably due to low transfection efficiency, low activity of the SV40 promoter, or a combination of both. Transfection of these cell types was repeated using different cationic lipid transfection reagents and five to ten times higher amounts of total RNA were loaded in the Northern blot experiments. To circumvent the possibility of low promoter activity, the SV40 promoter was replaced with the mouse beta adducin spleen specific one and its activity tested in MEL cells. However, no signal was detected after the introduction of the above-mentioned changes (data not shown). At this point, with the main aim of characterizing the cis-acting regulatory elements and trans-acting factors involved in the use of the distal PAS4, we decided to do this analysis in HeLa cells. First, we looked for the core cis-acting elements regulating cleavage and polyadenylation. Putative polydenylation core elements, hexanucleotide motif (Hm), upstream and downstream sequence elements (USE and DSE), were identified after visual inspection of the sequence and in silico comparison with the reported consensus sequences of these elements (Figure 17A, open boxes). To test their functionality, each of these putative elements was deleted in the original **A1-A23-A4** minigene resulting in the three mutant constructs: **A1-A23-A4[**ΔHm**]**, **A1-A23-A4[**ΔUSE**]** and **A1-A23-A4[**ΔDSE**]**. Deletion of the USE caused a decrease in the mRNA levels, probably due to a reduction in polyadenylation efficiency (Figure 17B). Absence of either the Hm or the DSE resulted in shorter mRNAs (Figure 17B). 3'RACE (Figure17C) and subsequent sequencing analysis of the PCR products showed that these differences in the size of the mRNA were due to the activation of two cryptic PASs, located about 200 bases upstream of the natural PAS4 (Figure 17A, indicated as "cryptic PAS4 (a) and (b)"). Therefore, these experiments confirmed the role and importance

of the core elements for the definition and correct 3'end processing of the beta-adducin pre-mRNA at the PAS4.

I should mention that the described study of the cis-acting elements had started prior to my arrival at the lab. I have contributed to this work studying the effect of the deletion of putative DSE from the PAS4. To this end, I have performed Northern blot, 3'RACE and sequencing analysis. I have also performed the transfection experiments of the **A1-A23-A4** construct in MEL, K562 and SK-N-BE cells to evaluate alternative PAS use.

After the determination of the cis-acting elements of the PAS4, this research line was directed towards the identification of trans-acting factors recognizing these cis-acting elements and participating in the 3'end processing of the beta-adducin pre-mRNA at the PAS4. These studies, and those related to the characterization of the core cis-acting elements mentioned above, were conducted by a postdoctoral fellow in the lab, Dr. Luisa Costessi, and are not part of this thesis. Instead, my research project focused on the characterization of the non-canonical cis-acting elements regulating polyadenylation of the beta-adducin pre-mRNA and the obtained results are described below.

2.1.1 Influence of the position of the PAS4 for its usage

To understand whether the absence of activity of the PAS1 and PAS23, and the use of the PAS4 site were influenced by their relative position, the additional **A4** region was inserted upstream of the two proximal PASs (Figure 18A, **A4-A1-A23-A4** construct). After transfecting this construct into HeLa cells, no signal was obtained by Northern blot analysis (Figure 18B, Lane 2). This result shows that neither the artificially "proximal" PAS4, nor the artificially "distal" PAS1 and PAS2-3 or the original distal PAS4 were used in this particular construct. Therefore, these results suggested that both the position and order of the different PASs present in the 3'UTR of the beta-adducin pre-mRNA may influence their usage.

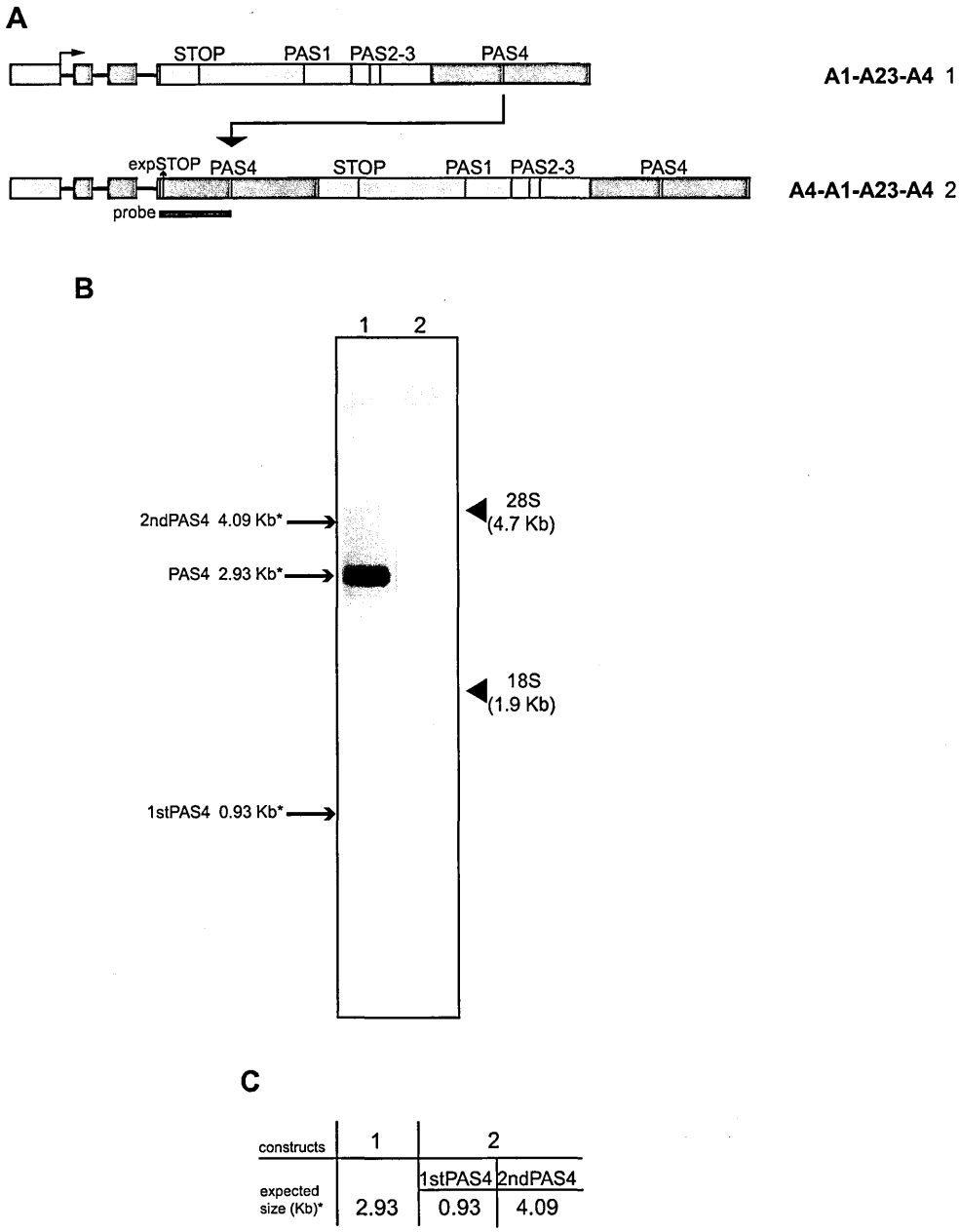
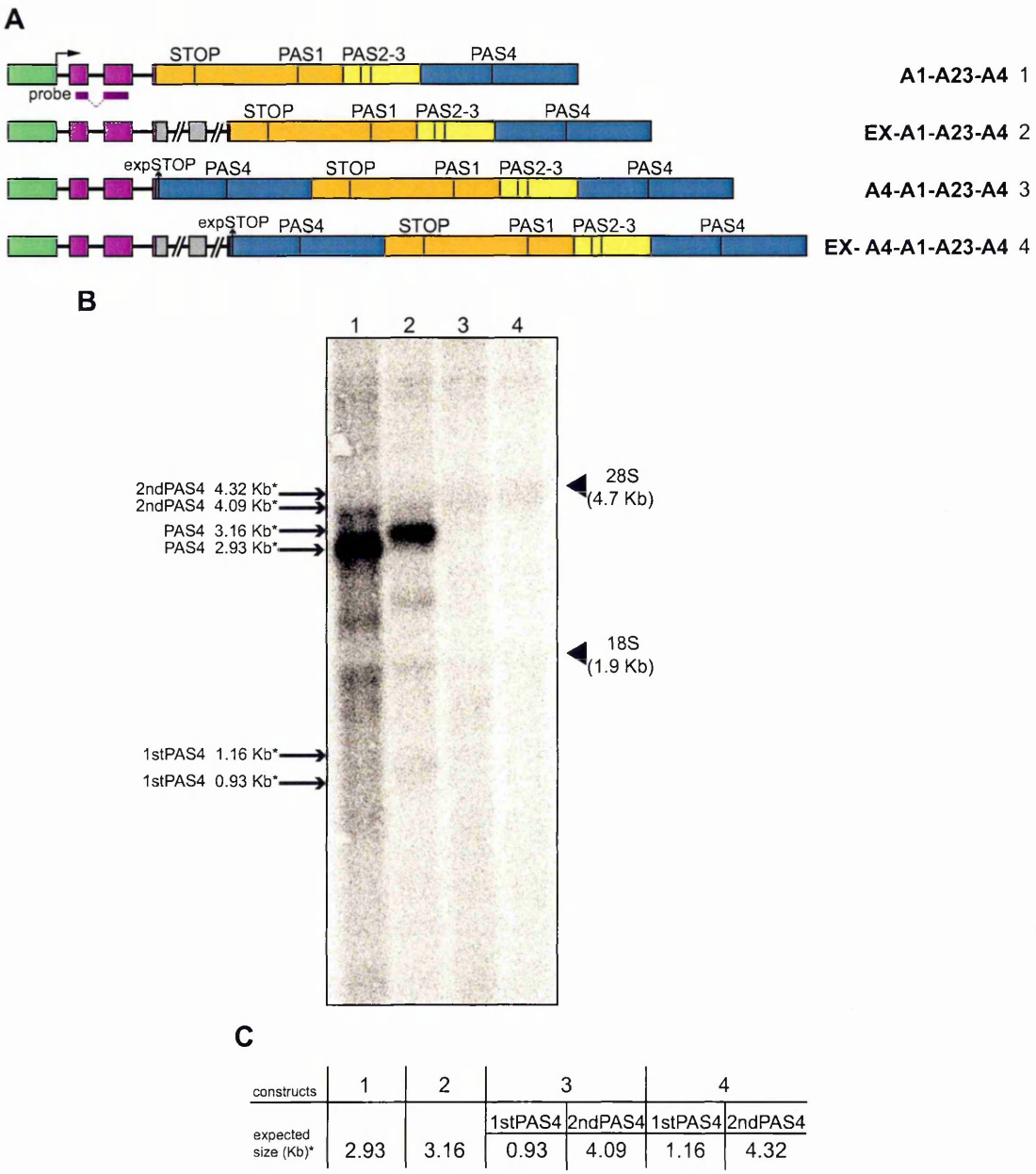


Figure 18
Influence of the position of the PAS4 for its usage.
A: schematic representation of the mouse beta adducin chimeric minigene constructs, **A1-A23-A4** (1) and **A4-A1-A23-A4** (2); STOP indicates the position of the beta adducin stop codon, which is also the expected stop codon of the globin-adducin chimera in the construct 1; expSTOP indicates the expected stop codon of the globin-adducin chimera in the construct 2. **B**: Northern blot analysis of 1,5 µg total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; there is no control for the transfection efficiency. **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4, (i.e., the asterisk indicates without the pA tail).



C

constructs	1	2	3		4	
			1stPAS4	2ndPAS4	1stPAS4	2ndPAS4
expected size (Kb)*	2.93	3.16	0.93	4.09	1.16	4.32

Figure 19

Effect of the addition of beta adducin introns on PAS4 usage.

A: schematic representation of the mouse beta adduci n chimeric minigene constructs, **A1-A23-A4** (1) **EX-A1-A23-A4** (2) and **A4-A1-A23-A4** (3) and **EX-A4-A1-A23-A4** (4); the alpha globin fragment used as a probe for the Northern blots is indicated; STOP indicates the position of the beta adducin stop codon, which is also the expected stop codon of the globin-adducin chimera in the construct 1 and 2; expSTOP indicates the expected stop codon of the globin-adducin chimera in the constructs 3 and 4. **B**: Northern blot analysis of 10 µg total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; there is no control for the transfection efficiency. **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail).

In order to generate more physiologically relevant minigenes, I replaced the alpha globin terminal exon with a fragment that included sequences from the mouse beta adducin Exon 14, 15 and the first 16 bases of exon 16 and the respective introns into the A1-A23-A4 and A4-A1-A23-A4 constructs (Figure 19A). The expression pattern determined by Northern blot was similar to that of minigenes not-containing the extra exons and introns (Figure 19B), indicating that the introduced sequences had no influence on pre-mRNA processing and PAS choice in the reporter transripts.

However, the absence of expression from both **A4-A1-A23-A4** constructs was very surprising. It appeared that the position of the PAS4 was important and that it was not processed when placed upstream of the proximal sites. But, what also remained unexplained was why the distal PAS4 in the **A4-A1-A23-A4** was not functional.

Once the pre-mRNA is cleaved at the PAS, termination of the transcription should occur somewhere in the region downstream of the cleavage site. Thus, we reasoned that the region downstream of the PAS4 could contain elements causing RNA Polymerase II to terminate or pause transcription. By inserting the PAS4 sequence in the upstream region in the **A4-A1-A23-A4** construct, premature termination or pausing could occur in these constructs. To test this hypothesis, I removed the 552 bp long region downstream of the "proximal" PAS4 DSE from both the **A4-A1-A23-A4** and **EX-A4-A1-A23-A4** constructs (Figure 20A, **A4[**5405-6015**]-A1-A23-A4** and **EX-A4[**5405-6015**]-A1-A23-A4** constructs). No signal was observed by Northern blot analysis of RNAs prepared from HeLa cells transfected with these constructs (Figure 20B), suggesting that this region was not responsible for the absence of any detectable transcript from the **A4-A1-A23-A4** and **EX-A4-A1-A23-A4** constructs.

Overall, we could conclude that:

- The position of the PAS4 was important
- The addition of the mouse beta-adducin exons 14 and 15, and their corresponding introns to improve the genomic context had no influence on the efficiency of the pre-mRNA processing and PAS choice.

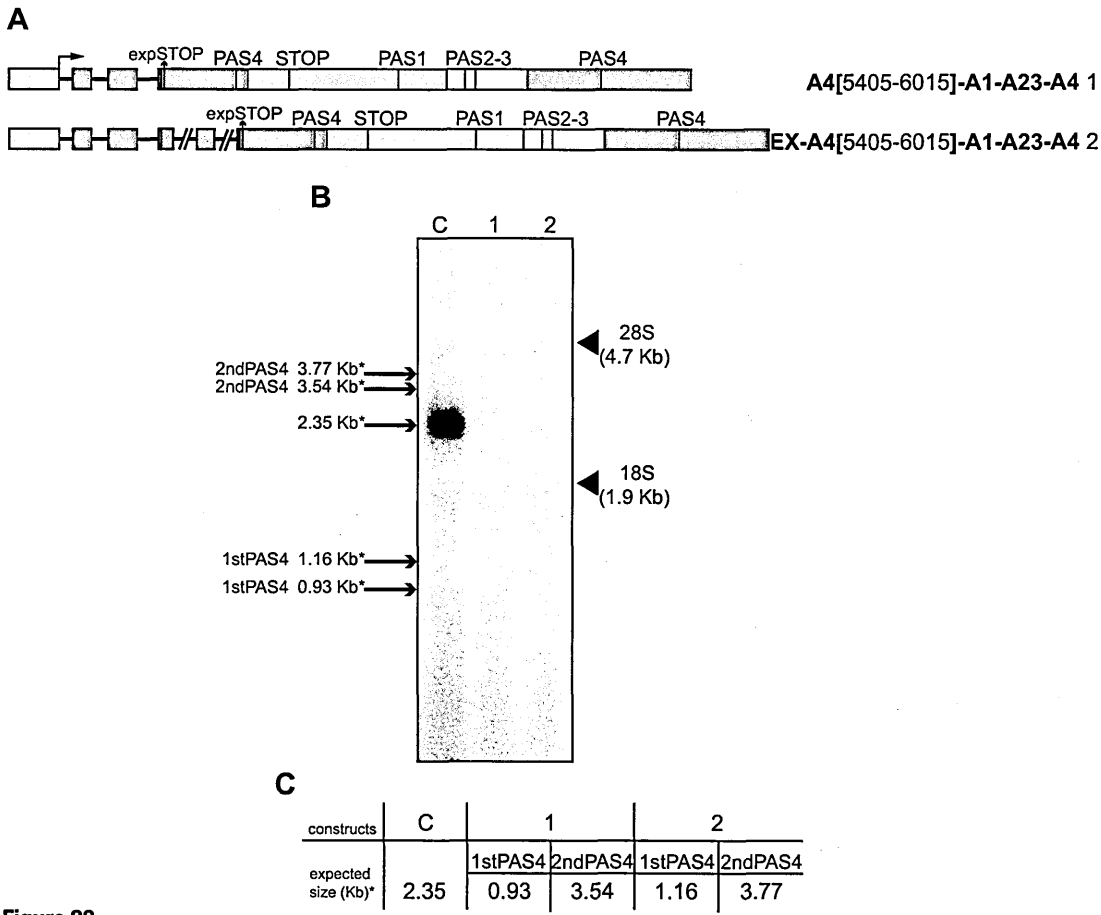


Figure 20

Influence of the position of the PAS4 for its usage: deletion of the downstream region of the proximal PAS4.

A: schematic representation of the mouse beta adducin chimeric minigene constructs, **A4[**5405-6015**]-A1-A23-A4** (1) and **EX-A4[**5405-6015**]-A1-A23-A4** (2); expSTOP indicates the expected stop codon of the globin-adducin chimera. **B**: Northern blot analysis of 5 µg total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; C - control RNA; there is no control for the transfection efficiency. **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail).

• The deletion of the sequences placed downstream of the proximal PAS4 (in the **A4[**5405-6015**]-A1-A23-A4** and **EX-A4[**5405-6015**]-A1-A23-A4** constructs) did not activate the use of the distal PAS4.

The precise reasons why the **A4-A1-A23-A4** and **EX-A4-A1-A23-A4** constructs showed no expression in the Northern blot still remains unclear.

It should be taken into account that the above-mentioned experiments were performed in the absence of transfection efficiency controls and thus conclusions drawn are only qualitative. The preliminary initial experiment with the **A4-A1-A23-A4** construct was performed by Dr. Luisa Costessi (Figure 18), prior to my arrival to the lab but this research line was temporarily postponed until my arrival to the lab. I have prepared the rest of the above mentioned constructs and performed experiments represented in Figures 19 and 20.

2.2 Identification of non-canonical cis-acting elements regulating usage of the distal PAS4 in the beta-adducin pre-mRNA

Above, I have described the studies that revealed the canonical cis-acting elements necessary for the correct definition and 3'end processing of the beta-adducin pre-mRNA at the PAS4 (Figure 17). These studies were conducted in a context where all three known PASs of beta-adducin were present. Considering the results we obtained with the **A4-A1-A23-A4** construct in the previous section, we wondered whether the presence of a proximal PASs could have an impact on the usage of the distal PAS4. To test this possibility, the regions containing the proximal PASs, **A1** and **A23**, were removed from the original minigene **A1-A23-A4** construct generating the **A4** construct (Figure 21A). Surprisingly, no signal was detected after transiently transfecting the **A4** construct into HeLa cells (Figure 21B, Lane 2). Adding back only the **A23** region (Figure 21A, construct **A23-A4**) was not enough to completely restore the mRNA levels detected for the **A1-A23-A4** construct (Figure 21B and D, Lane 3 compared to Lane 1). Interestingly, adding back only the **A1** region (Figure21A, construct **A1-A4**) restored the activity of the PAS4 (Figure 21B, Lane 4). In addition, the presence of the **A1** region immediately upstream of PAS4 increased mRNA levels by two fold when compared to the product obtained with the **A1-A23-A4** construct (Figure 21B and D, Lane 4 compared to Lane 1). On the contrary, a construct containing the **A1** region in the antisense orientation showed no activity (Figure 21A, construct **INA1-A4**, Figure 21B, Lane 5), ruling out that the increase in mRNA levels produced by the **A1-A4** construct could be related to the positioning of the PAS4 relative to the terminal exon. Consequently, the **A1** region contains cis-acting regulatory elements important to make transcripts viable.

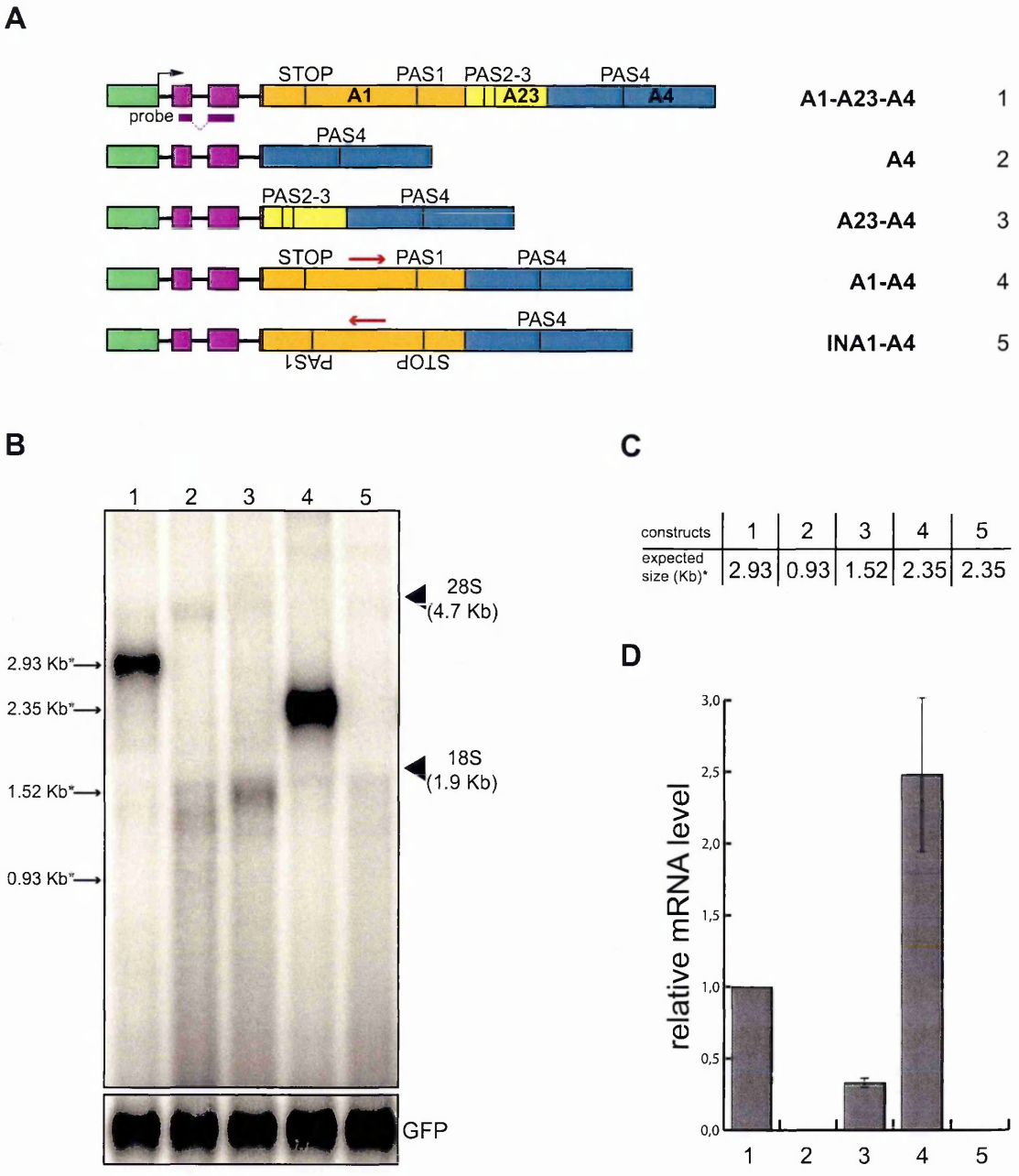


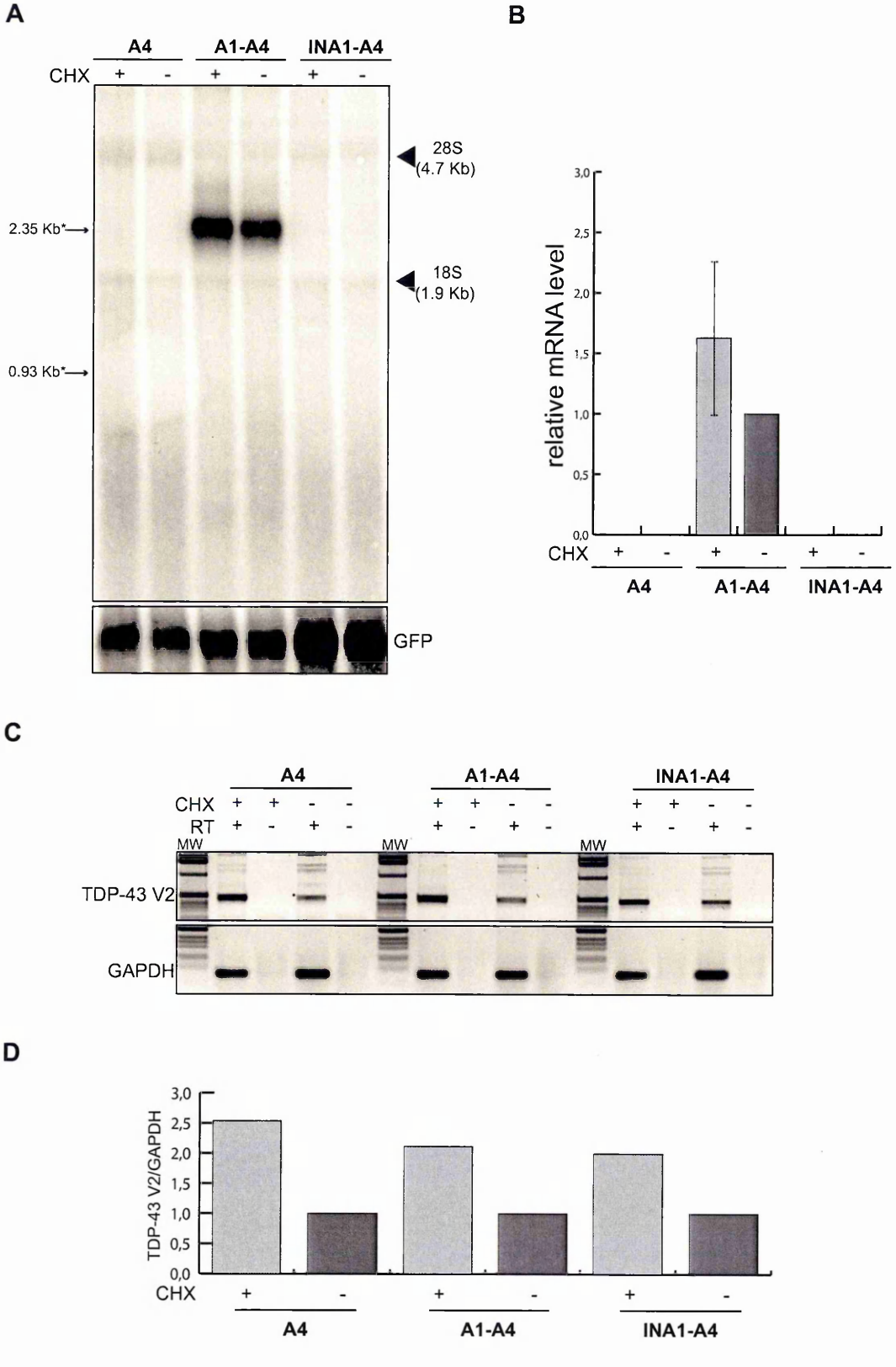
Figure 21

The presence of the A1 region is necessary to get chimeric minigene construct expression.

A:Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A4** (2), **A23-A4** (3), **A1-A4** (4) and **INA1-A4** (5); the alpha globin fragment used as a probe for the Northern blots is indicated; the green box indicates the SV40 promoter; the orange, yellow and blue boxes indicate **A1**, **A23** and **A4** region, respectively; the different beta adducin polyadenylation sites (PAS) and stop codon are indicated. **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). **D**: the relative mRNA levels were calculated as the ratio between the analyzed mRNAs and GFP mRNA represented in Panel B; quantification of the Northern blot signals has been performed using the Cyclone Phosphorimager and the OptiQuant software (Packard); the mean +/- S.D. of three independent experiments is shown.

One of the possible reasons for the absence of signal observed with the **A4** and **INA1-A4** constructs could be the degradation of the transcribed mRNA. mRNA degradation can be triggered as a part of mRNA surveillance pathway activated during translation (Garneau et al., 2007). If the mRNAs produced from the **A4** and **INA1-A4** constructs were degraded during translation, then blocking translation of mRNAs should lead to an increase in the levels of **A4-** and **INA1-A4**-derived mRNAs, should this become visible by the Northern blot. To prevent translation dependent mRNA degradation, after transfection of the **A4** and **INA1-A4** constructs, HeLa cells were treated with cycloheximide, an inhibitor of protein synthesis in eukaryotes. I have prepared total RNA from the cycloheximide-treated cells and analyzed it by Northern blot. As can be seen in Figure 22A, despite the inhibition of translation, no mRNAs were detected in RNA isolations from cells transfected with either **A4** or **INA1-A4** construct. As a control for the efficiency of the cycloheximide treatment, I have performed semiquantitative RT-PCR of a TDP-43 mRNA variant generated by alternative splicing, named V2, known to undergo nonsense mediated decay (NMD) (Figure 22C and D) (Ayala et al., 2011). Therefore, I can conclude that the absence of expression of the **A4** and **INA1-A4** constructs was not due to mRNA degradation by mRNA surveillance pathways activated during translation. In the nucleus, inefficiently polyadenylated mRNAs are targeted, retained near or at the site of their transcription and degraded by the nuclear RNA exosome complex (Hilleren et al., 2001; Schmid and Jensen, 2008). Therefore, another possible cause for the absence of expression from the **A4** and **INA1-A4** constructs could be inefficient 3'end processing of the pre-mRNA at the PAS4 which could lead to its degradation. Additionally, the observed differences in

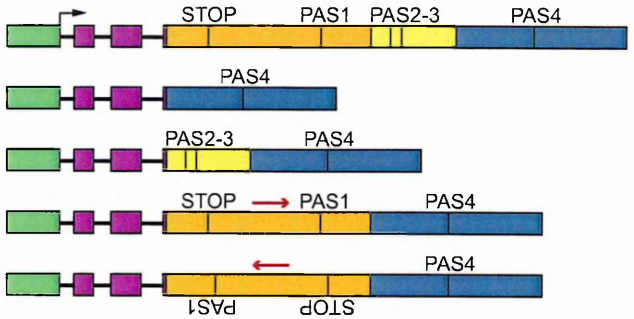
Figure 22 >>
Determination of RNA degradation by cycloheximide treatment.
A: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the **A4**, **A1-A4** and **INA1-A4** constructs; 24h after the transfection 50 µg/ml cycloheximide (CHX) was added to the cells, and after 3 hours cells were collected and RNA was prepared; the p-EFGP-C2 plasmid was used to normalize the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected mRNA sizes are indicated on the left; **B**: The relative mRNA levels were calculated as described for Figure 21; the mean +/- S.D. of three independent experiments is shown. **C**: RT-PCR analysis of TDP-43 V2 and GAPDH transcripts from the same RNAs used for the Northern blot shown in Panel A; RT indicates reverse transcriptase, MW indicates molecular weight marker; **D**: Quantification of the TDP-43 V2 normalized to the GAPDH signal obtained in Panel C; the quantification has been performed using Quantity One software (BioRad).


A
A4
A1-A4
INA1-A4
CHX + - + - + -
28S
(4.7 Kb)
18S
(1.9 Kb)
2.35 Kb*
0.93 Kb*
GFP
B
relative mRNA level
3,0
2,5
2,0
1,5
1,0
0,5
0,0
CHX + - + - + -
A4
A1-A4
INA1-A4
C
A4
A1-A4
INA1-A4
CHX + + - - + + - - + + - -
RT + - + - + - + - + - + -
MW
MW
MW
TDP-43 V2
GAPDH
D
TDP-43 V2/GAPDH
3,0
2,5
2,0
1,5
1,0
0,5
0,0
CHX + - + - + -
A4
A1-A4
INA1-A4

the expression levels of the constructs shown in Figure 21 could be the consequence of differences in the efficiency of 3'end processing at the PAS4. To test this hypothesis, I have performed RNase protection assays (RPA) with the same RNA isolations previously used for the Northern blot experiments shown in Figure 21. For the RPA, a 324 nt long input probe (IP) was used, that has 289 nt complementary to the pre-mRNA transcript (205nt upstream and 84nt downstream of the PAS4. Figure 23B). As a positive control, a total RNA from adult mouse brain (mBR) was also assayed. In addition to the expected band (205nt), two additional mBR protected products were observed (Figure 23C, asterisks). However, these products are likely unspecific as in the Northern blot and 3'RACE analyses only the expected product was detected (Figure 17B and C) and there is no EST-support for the presence of additional beta adducin PASs downstream of the PAS4. The cleaved protected product (pA) was clearly detected for the **A1-A23-A4**, **A23-A4** and **A1-A4** constructs (Figure 23C, Lanes mBR, 1, 3 and 4), but not for **A4** and **INA1-A4** (Figure 23C, Lanes 2 and 5), as already observed with the Northern blot analysis (Figure 21B). In addition, uncleaved protected products (RT-readthrough) were detected in all tested constructs at roughly the same levels, suggesting that all constructs were transcriptionally active. Cleavage efficiency was calculated as the ratio between the signal obtained for the pA product and that of the RT band, as described in Dalziel et al 2007. The **A1-A4** construct showed a twofold increase in the cleavage efficiency in comparison with the **A1-A23-A4** one (Figure 23D, Lane 4 compared with Lane 1). The **A23-A4** construct showed more than two-fold decrease in cleavage efficiency compared to the **A1-A23-A4** construct (Figure 23D, Lane 3 compared with Lane 1).

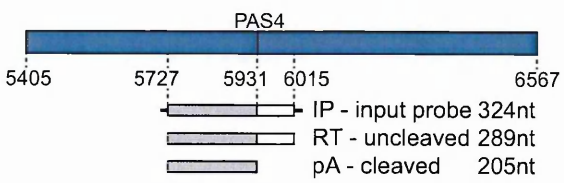
Figure 23 >>
Cleavage efficiency of the pre-mRNA is enhanced by the presence of the A1 region.
A: Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A4** (2), **A23-A4** (3), **A1-A4** (4) and **INA1-A4** (5). **B**: Schematic representation of the **A4** region containing PAS4 and the probe used in the RNase protection experiment shown below. The expected protected fragments, with their expected sizes are shown. Nucleotide positions are numbered respective to the first base of the last mouse beta adducin exon. **C**: RNase protection analysis of total RNA prepared from mouse brain (mBR) and HeLa cells previously transfected with the constructs shown in panel A (1-5); IP, input probe; RT, uncleaved protected product; pA, cleaved protected product; MWM, molecular weight marker; the asterisk indicates unidentified products in the mBR sample. **D**: The cleavage efficiency was calculated as the ratio between cleaved and uncleaved protected products represented in Panel C (1-5). The quantification of the protected signals has been performed using the Cyclone Phosphorimager and the OptiQuant software (Packard); the mean +/- S.D. of three independent experiments is shown; data were analyzed using one-way ANOVA and Tukey's multiple comparison test, *** $p<0.0001$.

A

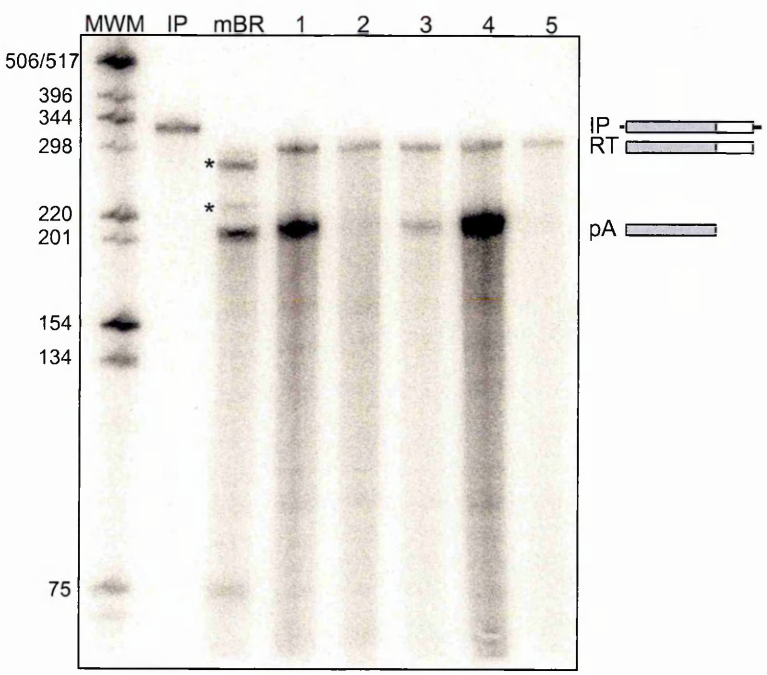


A1-A23-A4	1
A4	2
A23-A4	3
A1-A4	4
INA1-A4	5

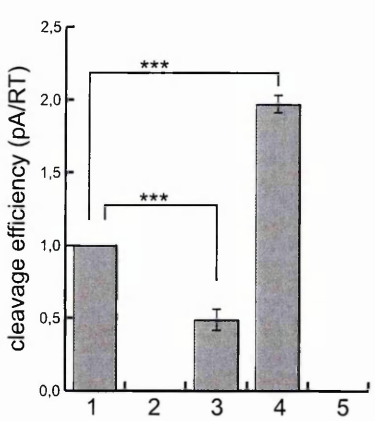
B



C



D



Interestingly, there is a similarity between the observed differences in the cleavage efficiency and the amount of mRNAs detected by the Northern blot analysis (compare Figure 23D with Figure 21D). These results support the hypothesis that the differences in mRNA expression levels from the described constructs may be caused by a reduction in 3'end processing efficiency.

The readthrough products in all tested RNA isolations, are roughly at similar levels, suggesting that all constructs are transcriptionally active. RT-qPCR amplification of the pre-mRNA, with primers annealing to sequences in the first intron and the second exon of the alpha globin gene, respectively, showed similar levels of pre-mRNA for constructs **A4**, **A23-A4**, **A1-A4** and **INA1-A4** which were roughly half of **A1-A23-A4** pre-mRNA levels (Figure 24).

Summarizing the above-mentioned results, I could conclude that:

- the **A1** and **A23** regions contain cis-acting elements which are required for cleavage and polyadenylation of the beta-adducin pre-mRNA at the PAS4. Deletion of the **A1** and **A23** regions from the **A1-A23-A4** construct results in the absence of expression (**A4** construct).
- pre-mRNA processing activity at the PAS4 is rescued by the addition of the **A1** region upstream of the **A4** segment (**A1-A4** construct), as this construct is more active than the **A1-A23-A4** construct. The **A1** region appears to contain positive cis-acting element(s) enhancing 3'end processing at the PAS4.
- the rescue in the activity of the PAS4 caused by the presence of **A1** region in the **A1-A4** construct was not related to the distance between the 3' splice site and the PAS4, as the same **A1** region in the inverted orientation led to the complete absence of the activity of PAS4.
- the **A23** region contains negative cis-element(s) suppressing 3'end processing at the PAS4, since its inclusion into the **A1-A4** construct (the **A1-A23-A4** construct) led to a significant decrease of **A1-A23-A4** mRNA levels, possibly by decreasing the cleavage efficiency at the PAS4.

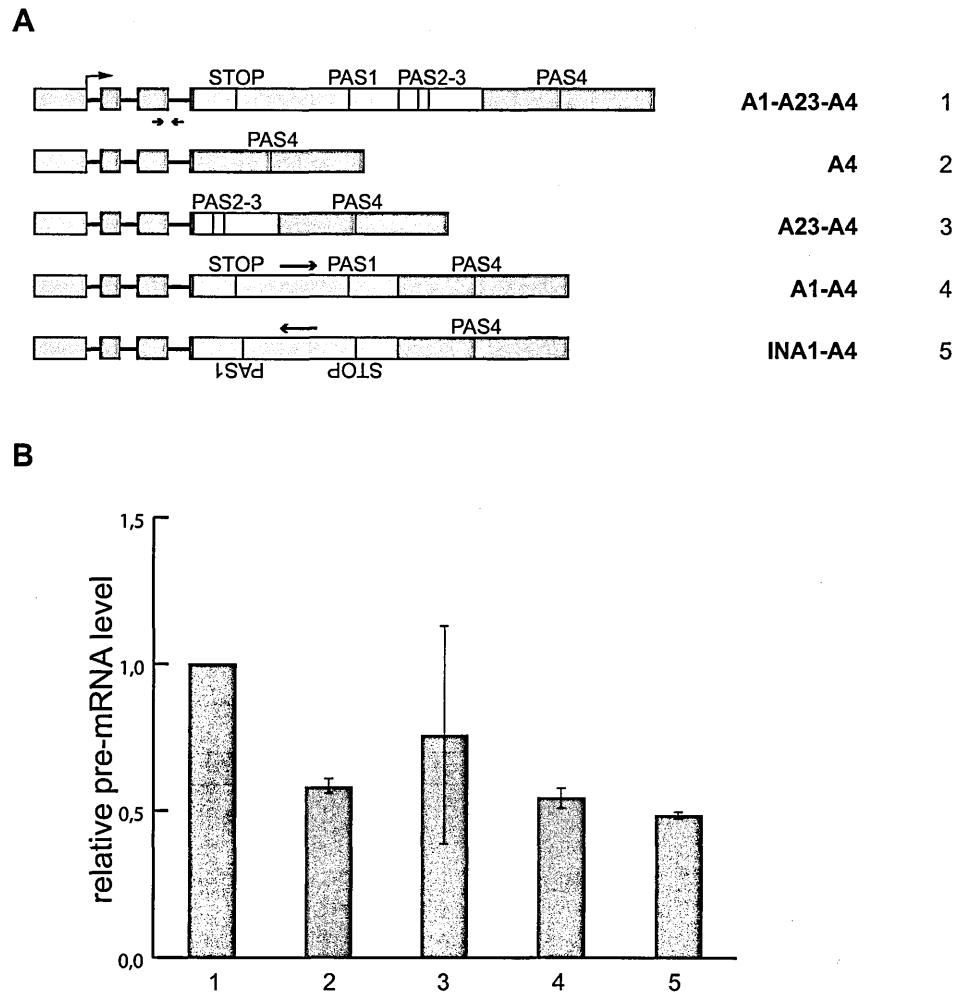


Figure 24
Determination of pre-mRNA levels by RT-qPCR.
A: Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A4** (2), **A23-A4** (3), **A1-A4** (4) and **INA1-A4** (5); green arrows indicate primer pair used for the RT-qPCR showed in panel B. **B**: Quantification of the pre-mRNA levels by RT-qPCR from the total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the p-EFGP-C2 plasmid was used to normalize the transfection efficiency; the relative pre-mRNA levels are represented, measured as the ratio between the analyzed RNAs and the GFP control; the mean +/- S.D. of two independent experiments is shown; experiments were performed in duplicates.

Taken together, these results demonstrated the presence of novel non-canonical 3'end processing-regulatory elements in the **A1** and **A23** regions of the beta-adducin pre-mRNA that were capable of enhancing or suppressing, 3'end processing at the distal PAS4. The next series of studies were directed to the fine mapping of the enhancer and silencer elements present in the **A1** and **A23** regions.

2.2.1 Mapping of the silencer elements of the A23 region

Deletion of the **A23** region in the **A1-A23-A4** construct resulting in the **A1-A4** plasmid led to an important increase in the levels of expressed mRNA (Figure 21B and D, compare Lane 4 to 1; and Figure 23C, compare Lane 1 to Lane 4). Since all other possible mechanisms were ruled out (see above Figures 22 and 31), these results strongly suggest that the effects could be related to the presence of a negative cis-acting element within the **A23** region, which suppresses cleavage efficiency at the PAS4. In order to map more in detail this potential negative cis-acting element, a "deletion" strategy similar to that one used for the **A1** region was adopted. Deleted segments of the **A23** region were introduced at the place of the full size **A23** present in the **A1-A23-A4** construct (Figure 25A, constructs **A1-A23[**1416-1781**]-A4** and **A1-A23[**1416-1581**]-A4**). The expression profiles of the new constructs were determined by Northern blot analysis of total RNA prepared from transiently transfected HeLa cells. Transfection efficiency was normalized by cotransfection with a control plasmid expressing Renilla luciferase. Removal of the negative cis-acting element should be seen as an increase in the mRNA levels in the Northern blot, similar to that one observed after deletion of the whole **A23** region (Figures 21B and 23C, compare Lane 1 to Lane 4). I observed such an increase for the **A1-A23[**1416-1581**]-A4** construct (Figure 25B and D, compare Lanes 1 and 3). The observed increase was even bigger in comparison to the increase observed when whole **A23** is absent (Figure 21B and D, Lane 1 and 4). This result suggests the location of the silencer element within the 200 bp region that spans from 1581 to 1781 bp relative to the first base of the last exon of the mouse beta-adducin gene.

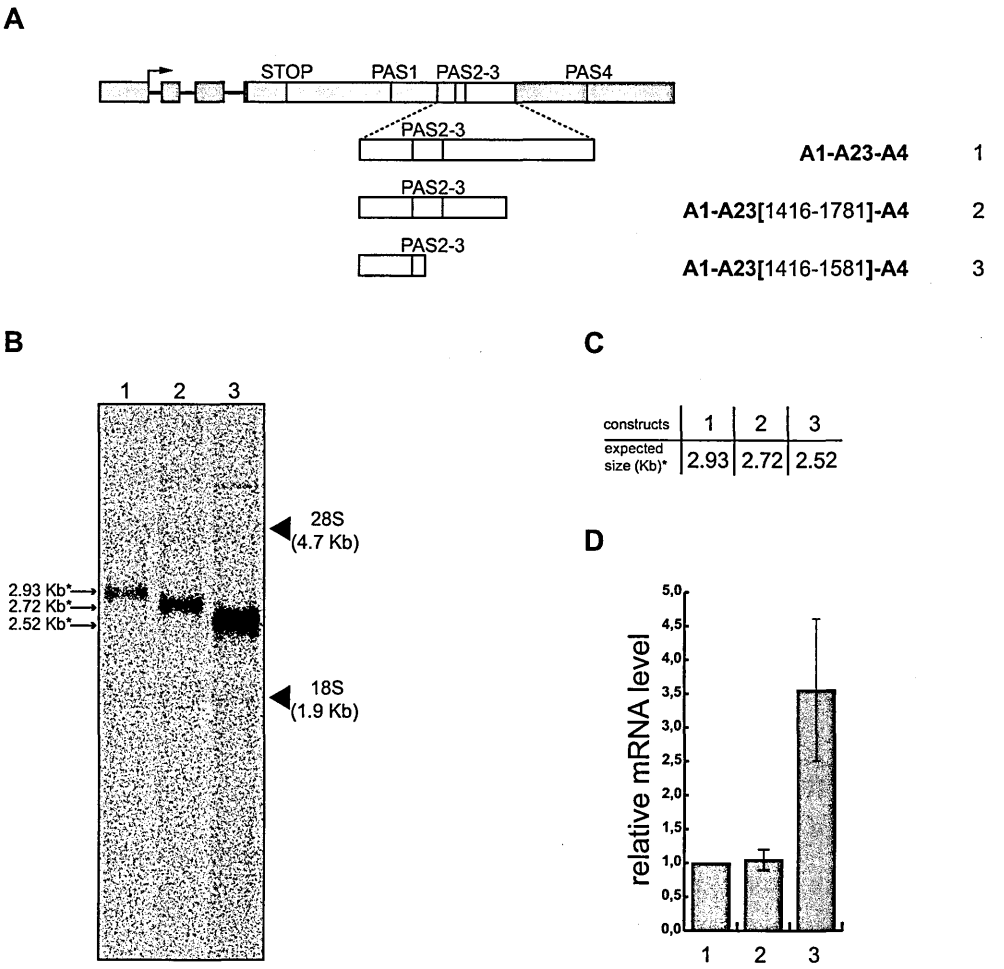


Figure 25
Mapping of the silencer element.
A:Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A1-A23[**1416-1781**]-A4** (2) and **A1-A23[**1416-1581**]-A4** (3). **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the alpha globin probe was used to detect the mRNAs of interest; phRG-TK (Renilla) plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). **D**: RT-qPCR analysis of the relative mRNA levels calculated as the ratio between analyzed mRNAs and Renilla control mRNA; the mean +/- S.D. of three independent experiments is shown; experiments were performed in duplicates.

2.2.2 Mapping of the enhancer elements present in the A1 region

The **A1** region spans 1400 bp comprising coding and noncoding sequences of the last mouse beta-adducin exon (Figure 16A). To have a deeper insight in the elements enhancing 3'end processing at the PAS4, the **A1** region was divided in three partially overlapping sub-regions, **A1[**17-305**]**, **A1[**284-887**]** and **A1[**868-1416**]**. Each of these sub-regions replaced the full **A1** region of the **A1-A4** construct, thus creating **A1[**17-305**]-A4**, **A1[**284-887**]-A4** and **A1[**868-1416**]-A4** constructs (Figure 26A). All constructs contained the intact **A4** region. The expression pattern of the new constructs was determined after transfection of HeLa cells followed by Northern blot analysis. Only the **A1[**17-305**]-A4** construct (containing the most proximal sub-region) produced a signal visible by Northern blot analysis, but the intensity of this signal was very faint (Figure 26B, Lane 2). Complete absence of the signal was observed for the other two constructs (Figure 26B, Lanes 3 and 4). These results suggest that the enhancer elements may be disrupted in those constructs. To identify the enhancer elements that were needed to recover the full function of the PAS4, I joined the **A1[**17-305**]** to the **A1[**284-887**]** sub-region, as well as the **A1[**284-887**]** to the **A1[**868-1416**]**. Next, I replaced the **A1** region upstream of the **A4** region by the **A1[**17-887**]** and **A1[**284-1416**]** segments to create **A1[**17-887**]-A4** and **A1[**284-1416**]-A4** constructs (Figure 26A). The presence of the **A1[**17-887**]** segment completely restored the signal obtained with the full **A1** region, whereas the **A1[**284-1416**]-A4** construct, which contained sequences downstream to the stop codon, produced no signal (Figure 26B, Lanes 5 and 6, respectively). Consequently, the **A1[**868-1416**]** segment containing the PAS1 had no influence on the processing at the PAS4. This result suggested that regions upstream and downstream of the stop codon contain elements that interact to enhance the activity of the PAS4. Next, I started with a more detailed analysis of the bipartite enhancer element present in the **A1** region.

In order to map the enhancer element located upstream of the stop codon, I have adopted an approach based on serial deletions starting from the 5'end of the **A1** region,

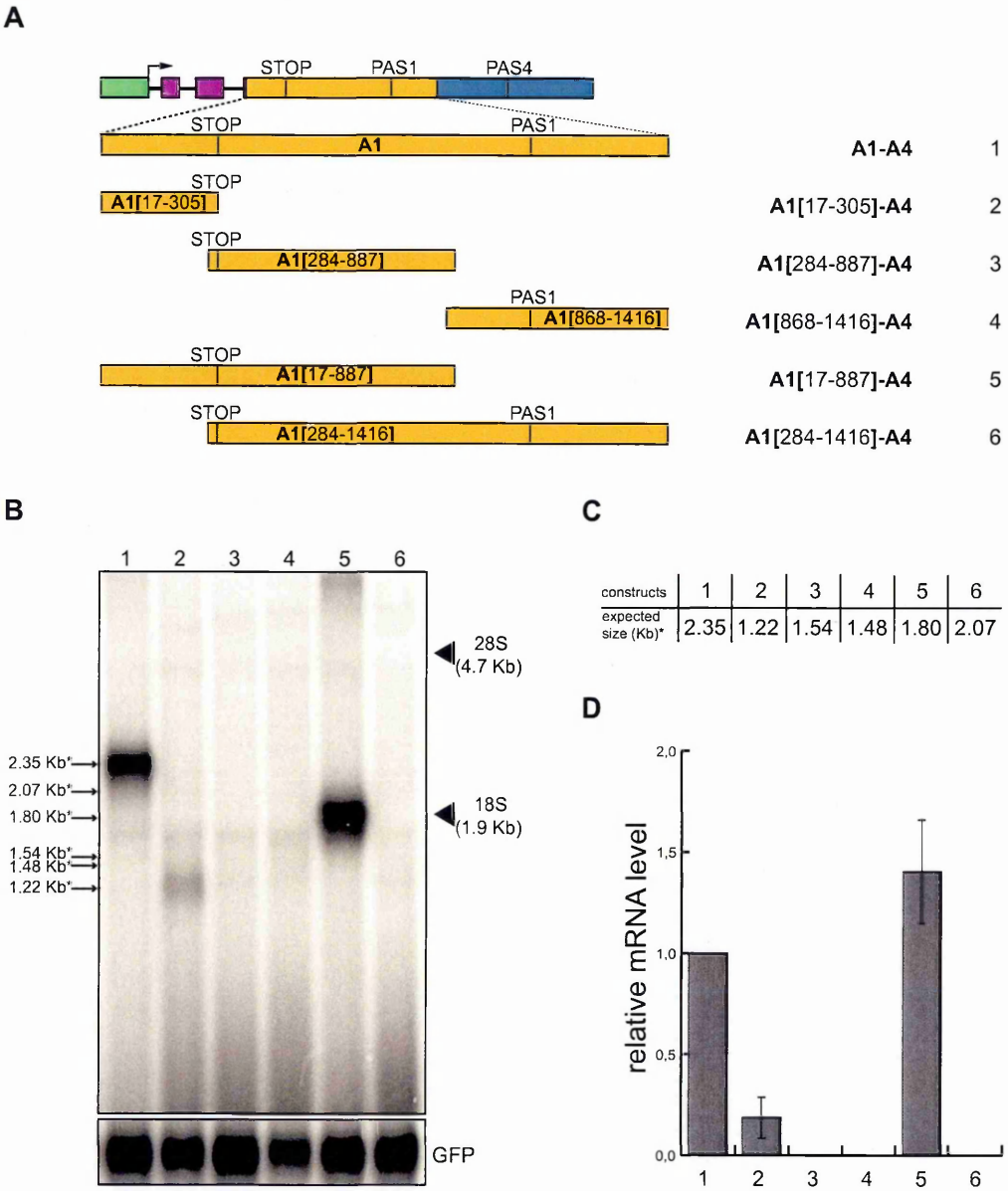


Figure 26

Mapping of the enhancer element present in the A1 region.

A:Schematic representation of the chimeric minigene constructs: **A1-A4** (1), **A1**[17-305]**-A4** (2), **A1**[284-887]**-A4** (3), **A1**[868-1416]**-A4** (4), **A1**[17-887]-A4 (5) and **A1**[284-1416]**-A4** (6). **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the alpha globin fragment used as a probe for the Northern blots is indicated in figure 21; the p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). D: the relative mRNA levels were calculated as described for Figure 21; the mean +/- S.D. of three independent experiments is shown.

thus partially eliminating the coding portion of the **A1** (Figure 27A). All deletions were done without altering the reading frame of the coding sequence. I have replaced the **A1** region from the **A1-A4** construct by these new deleted segments (Figure 27A). I expected that by deleting the specific bases of the enhancer element, the levels of the minigene-derived transcript should be affected. The difference in mRNA levels should be observed as an important reduction or complete absence of the signal in the Northern blot. All constructs tested generated the expected bands, except for the **A1[**284-1416**]-A4** construct (Figure 27B, Lane 4) where no band was observed. The only difference between the **A1[**260-1416**]-A4** construct, which still maintained full pre-mRNA processing activity, and the **A1[**284-1416**]-A4** construct, which was inactive, consisted in the presence of 24 nt in the **A1[**260-1416**]-A4** construct (Figure 27B, Lane 3). Therefore, the presence of these 24 nt were essential to confer pre-mRNA processing activity and were thus considered to be an upstream enhancer element.

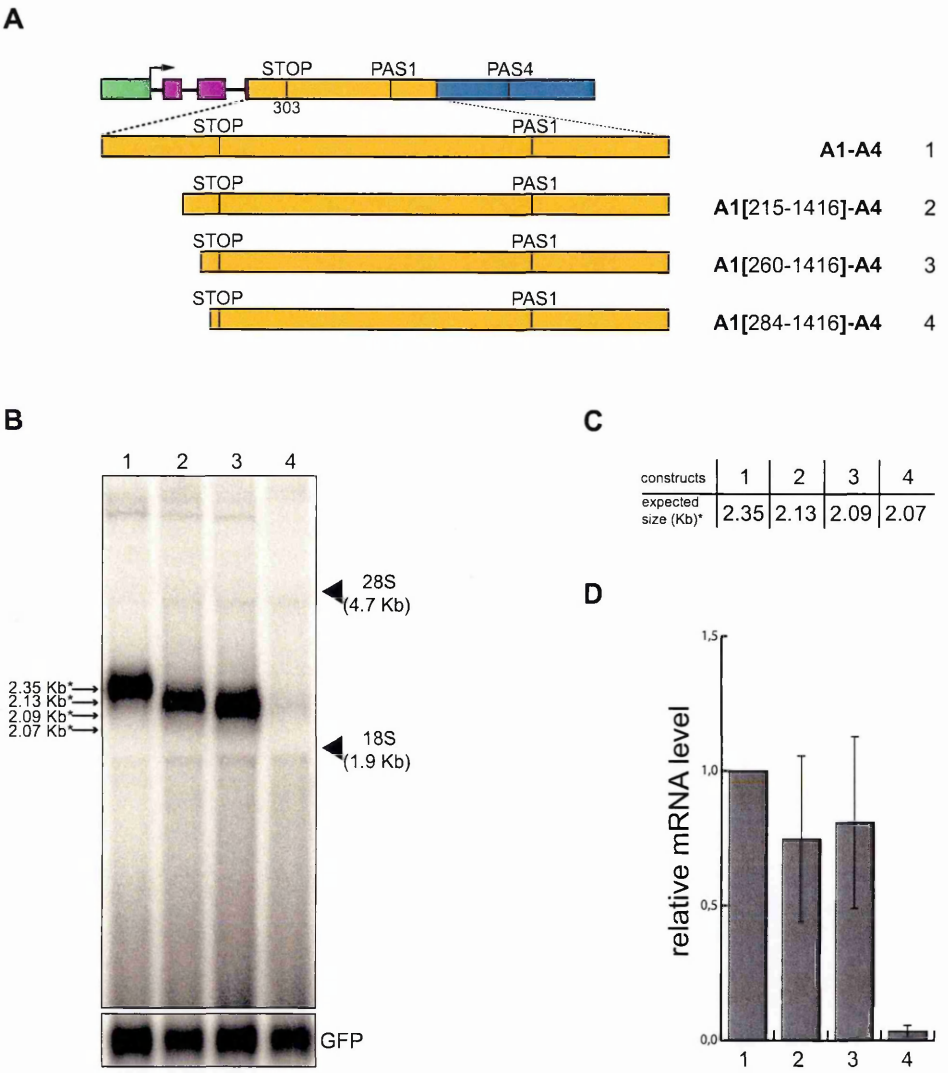


Figure 27

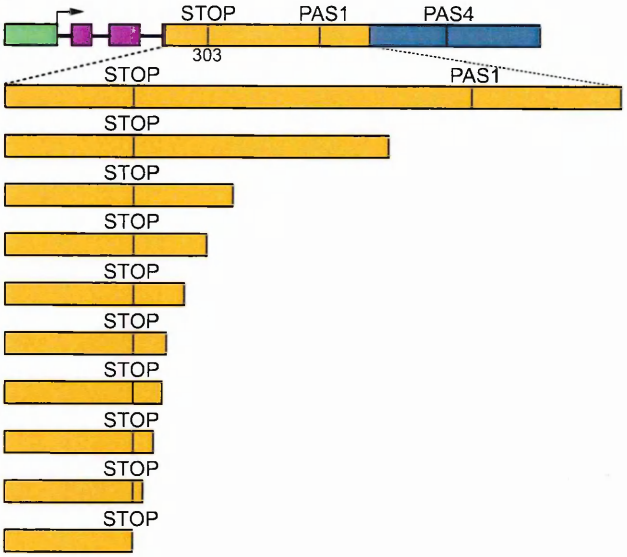
Mapping of the enhancer element present in the A1 region: upstream of the stop codon.

A:Schematic representation of the chimeric minigene constructs: **A1-A4** (1), **A1**[215-1416]**-A4** (2), **A1**[260-1416]**-A4** (3) and **A1**[284-1416]**-A4** (4). **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the alpha globin fragment used as a probe for the Northern blots is indicated in figure 21; the p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). **D**: the relative mRNA levels were calculated as described for Figure 21; the mean +/- S.D. of three independent experiments is shown.

To complement the above described approach, I have created serial deletions starting from the 3'end of the **A1** region. The fragments containing these deletions substituted the whole **A1** region present in the **A1-A4** construct (Figure 28A). All constructs contained the full **A4** region. As done previously, the effect of the deletions on mRNA expression was monitored by Northern blot. As expected, the reduction in the size of the detected mRNAs correlated, in each construct, with the length of the deletion (Figure 28B). Similarly to the approach adopted above, a significant drop of mRNA levels was observed when the deletion reached a region immediately downstream of the stop codon. All constructs generated a strong band in the Northern blot, except for the faint signal observed for the **A1[**17-305**]-A4** construct (Figure 28B, Lane 10). The difference between the constructs **A1[**17-305**]-A4** and **A1[**17-328**]-A4** was the presence of just 23bp in the latter. Therefore, these 23 nt in the **A1[**17-328**]-A4** construct were necessary to enable pre-mRNA 3'end processing activity at the PAS4 and were considered as the downstream enhancer element.

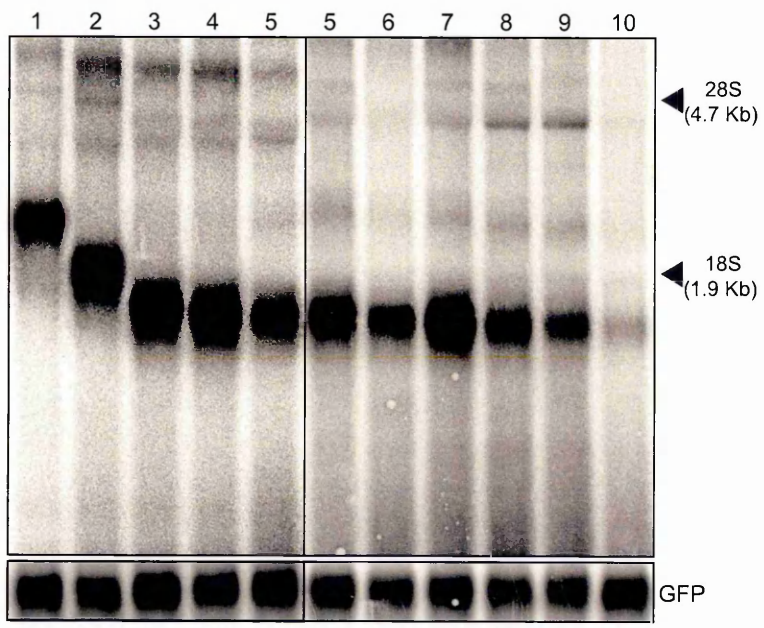
Figure 28>>
Mapping of the enhancer element present in the A1 region: downstream of the stop codon.
A:Schematic representation of the chimeric minigene constructs: **A1-A4** (1), **A1[**17-887**]-A4** (2), **A1[**17-531**]-A4** (3), **A1[**17-471**]-A4** (4), **A1[**17-420**]-A4** (5), **A1[**17-380**]-A4** (6), **A1[**17-371**]-A4** (7), **A1[**17-351**]-A4** (8), **A1[**17-328**]-A4** (9) and **A1[**17-305**]-A4** (10). **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in panel A; the alpha globin fragment used as a probe for the Northern blots is indicated in Figure 21; the p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right. **C**: the expected sizes of the mRNAs of the constructs shown in Panel A, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). **D**: the relative mRNA levels were calculated as described for Figure 21; the mean +/- S.D. of three independent experiments is shown; data were analyzed using one-way ANOVA and Tukey's multiple comparison test, *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ns - not significant for the $p<0.05$ as a threshold.

A



A1-A4	1
A1[17-887]**-A4**	2
A1[17-531]**-A4**	3
A1[17-471]**-A4**	4
A1[17-420]**-A4**	5
A1[17-380]**-A4**	6
A1[17-371]**-A4**	7
A1[17-351]**-A4**	8
A1[17-328]**-A4**	9
A1[17-305]**-A4**	10

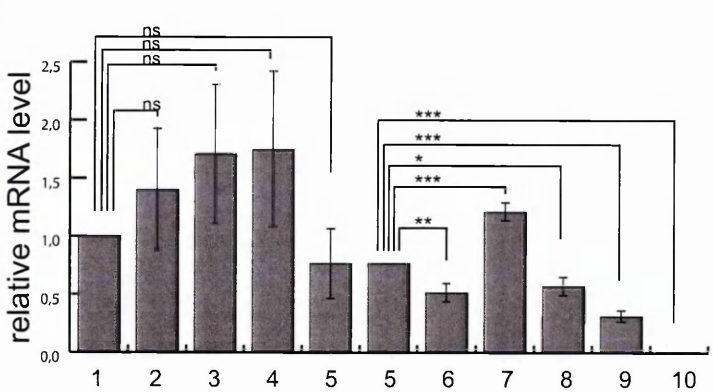
B



C

constructs	1	2	3	4	5	5	6	7	8	9	10
expected size (Kb)*	2.35	1.80	1.45	1.39	1.34	1.34	1.30	1.29	1.27	1.24	1.22

D



Taking both analyses together, it was possible to denote a composite enhancer element composed of at least of 69 bases, centered around the stop codon of the beta-adducin gene (Figure 29A). To test the functionality of the putative 69 bases-long enhancer element, I have made another set of constructs, by introducing deletions of either the upstream or downstream element, both together or the deletion of all 69 bases-long enhancer element into the **A1-A4** construct (Figure 29B). All of the deletion constructs showed the same level of expression, which was also similar to that of the **A1-A4** construct (Figure 29C and E). These results suggested that the 69 bases-long enhancer may represent an important but not essential element, depending on the context, to confer processing activity at the PAS4. Thus, other cis-acting elements located nearby could compensate for the absence of the 69 bases-long enhancer element, indicating redundancy of the system. However, the possibility remains that the 69 bases-long enhancer element may be enough to enhance the processing at the PAS4. To check for this possibility, I placed the 69 bases-long enhancer element upstream of the A4 region generating the **A1[**260-328**]-A4** construct (Figure 30A). A very faint signal was observed in the Northern blot analysis, after transfecting the construct into HeLa cells (Figure 30B, Lane 3). This results suggested that the 69 bases-long enhancer was not enough to activate processing of the **A1[**260-328**]-A4** transcript, and that additional elements are needed to provide full pre-mRNA processing activity full activity to the enhancer element. Therefore, in order to identify these additional elements, I placed a 166 bp segment upstream of the **A4** region, which contained about 50 bp of additional sequences at each side of the 69 bp element (Figure 30A, construct **A1[**215-380**]-A4**). The 166 bases-long segment was able to enhance the processing at the PAS4 and to rescue the signal in the Northern blot (Figure 30B, Lane 4). However, the detected signal did not reach the levels of the signal obtained from the **A1-A4** construct (Figure 30C, Lane 4 compared with 1), suggesting that other cis-acting sequences may still be necessary. Another possible explanation could be associated to differences in stability of the mRNAs. Therefore, I have compared

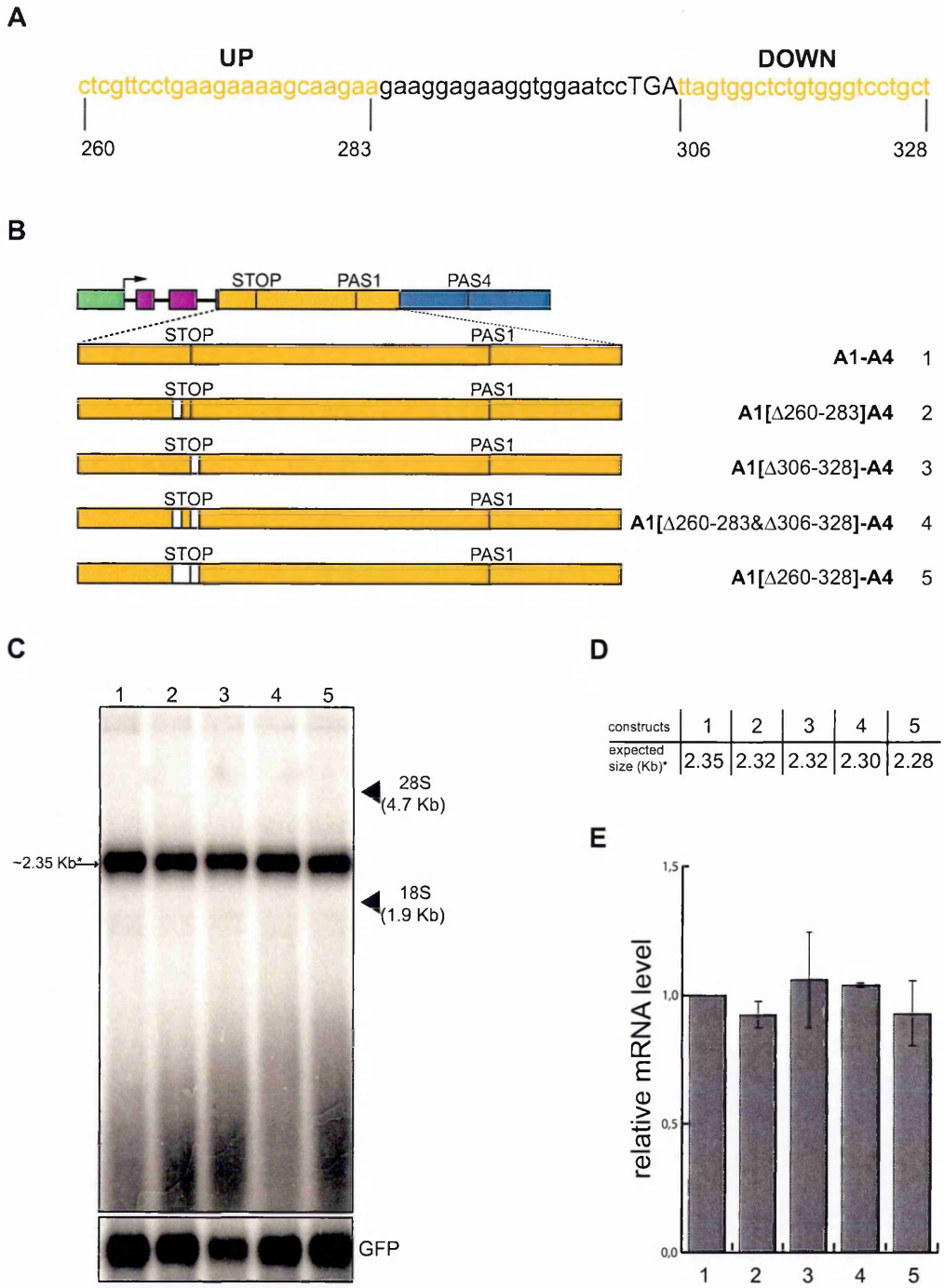


Figure 29

Deletion of the putative enhancer elements.

A: Sequence composition of the 69 nt long enhancer element; capital letters indicate the stop codon (TGA); UP and DOWN indicate the upstream and downstream elements, marked by orange-coloured letters; numbers correspond to the nucleotide position respective to the first base of the last mouse beta adducin exon. **B**: Schematic representation of the chimeric minigene constructs: **A1-A4** (1), **A1**[Δ260-283]**-A4** (2), **A1**[Δ306-328]**-A4** (3), **A1**[Δ260-283&Δ306-328]**-A4** (4) and **A1**[Δ260-328]**-A4** (5); the white boxes indicate the deleted regions. **C**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in Panel B; the alpha globin fragment used as a probe for the Northern blots is indicated in figure 21; p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right. **D**: the expected sizes of the mRNAs of the constructs shown in Panel B, up to the predicted PAS4 (i.e., the asterisk indicates without the pA tail). **E**: The relative mRNA levels were calculated as the ratio between analyzed mRNAs and GFP control mRNA shown in Panel C; quantification of the Northern blot signals has been performed as described above; the mean +/- S.D. of three independent experiments is shown.

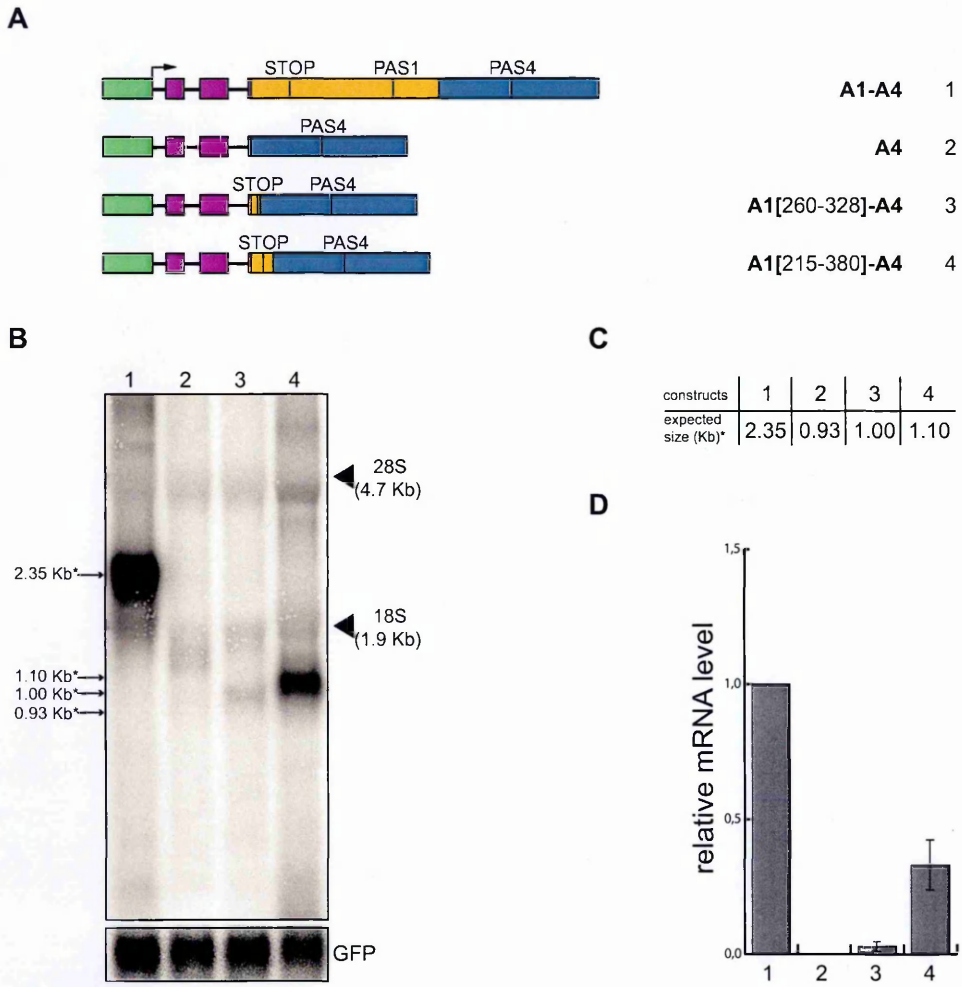


Figure 30
166nt long enhancer element is enough to activate the PAS4.
A: Schematic representation of the chimeric minigene constructs: **A1-A4** (1), **A4** (2), **A1[**260-328**]-A4** (3), **A1[**215-380**]-A4** (4); **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in Panel A; the alpha globin fragment used as a probe for the Northern blots is indicated in figure 21; p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected sizes shown in panel C are indicated on the left; **C**: the expected sizes of the mRNAs of the constructs shown in panel A up to the predicted PAS4, without pA tail (indicated by an asterisk). **D**: The relative mRNA levels calculated as the ratio between analyzed mRNAs and GFP mRNA control represented in panel B; quantification of the Northern blot signals has been performed as described above; the mean +/- S.D. of four independent experiments is shown.

mRNA stabilities for the **A1[**215-380**]-A4**, **A1-A4** and **A1-A23-A4** constructs.

In order to block the new synthesis of mRNAs and to monitor the decay of the already-synthesized ones, transfected HeLa cells were treated with actinomycin D and total RNA was collected at different time points. Cells were co-transfected with a plasmid coding for GFP to normalize for transfection efficiency and the amount of mRNA at the time zero was chosen as reference. As expected, the detected signal decreased with the Actinomycin D treatment, but the rate of decay was similar for all three mRNAs (Figure 31A and B) suggesting that there

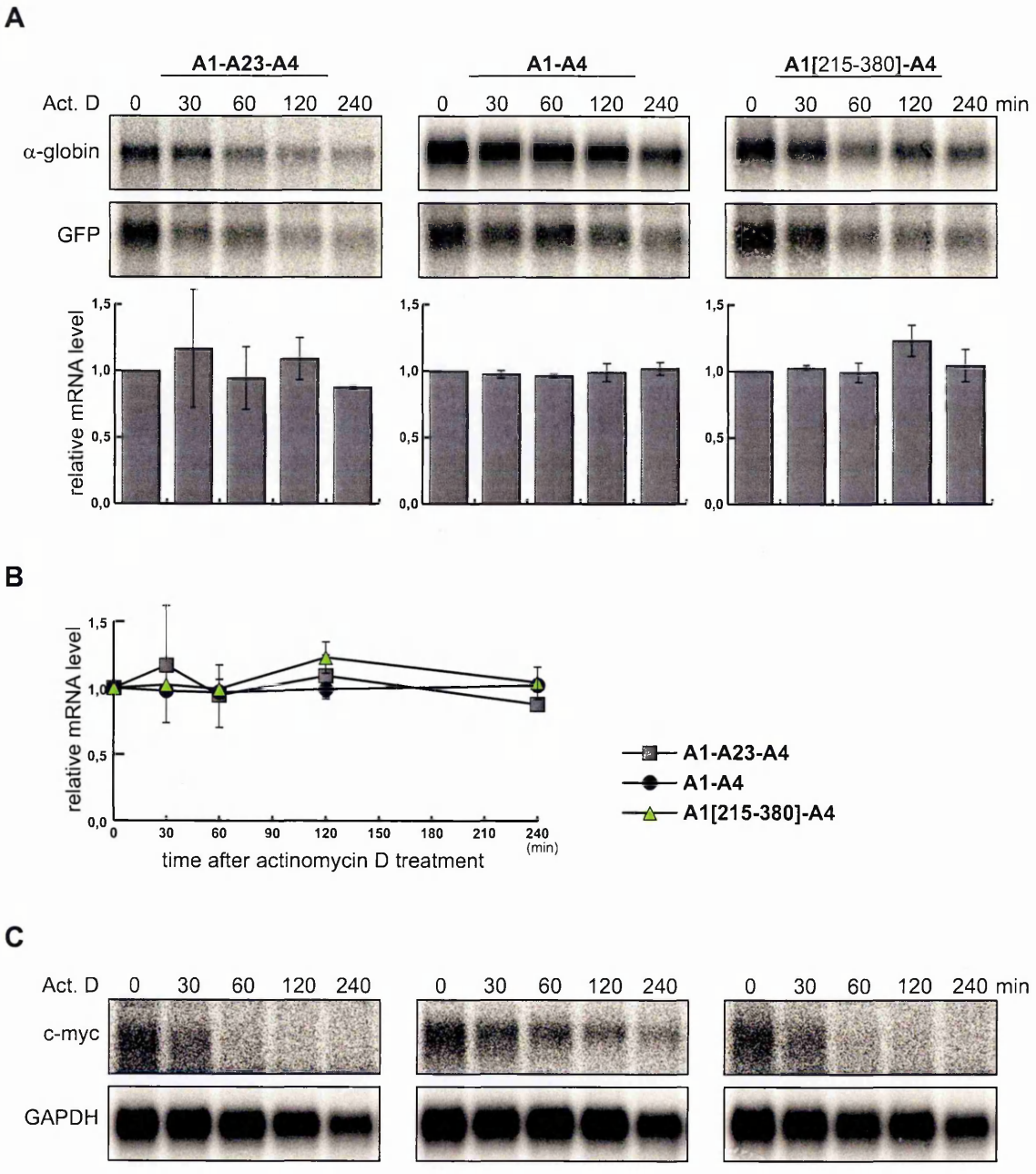


Figure 31

Determination of mRNA stability by actinomycin D treatment.

A: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the **A1-A23-A4**, **A1-A4** and **A1**[215-380]**-A4** constructs; 24h after the transfection 50 µg/ml actinomycin D (Act. D) was added to the cells; RNAs were collected at different time points after the Act. D treatment; the alpha globin probe was used to detect the mRNAs of interest; p-EGFP-C2 plasmid was used to normalize the transfection efficiency; the lower panel shows the relative mRNA levels calculated as the ratio between analyzed mRNAs and GFP control mRNA; quantification of the Northern blot signals has been performed using OptiQuant; mean +/- S.D. of the two independent experiments is shown; **B**: Comparison of the mRNA stability between **A1-A23-A4**, **A1-A4** and **A1**[215-380]**-A4** mRNAs. **C**: Northern blot analysis of the same RNAs shown in panel A, using probes to detect endogenous c-myc and GAPDH mRNA levels.

were no significant differences in the stability of the analyzed mRNAs. As a control for the efficiency of the actinomycin D treatment, I have analyzed the c-myc mRNA, which was reported to have a short half life, and the more stable GAPDH mRNA (Figure 31C) (Dani et al., 1984).

2.2.2.1 Initial characterization of possible trans-acting factors recognizing the enhancer element

After manual inspection of the 166 bases-long enhancer element (the full sequence is shown in Figure 32A), I observed that the region upstream of the stop codon is particularly rich in purines. It is known that purine-rich regions in pre-mRNAs are potential landing pads for the RNA binding proteins of the SR family, such as SRSF1 (former name SF2/ASF). This family of proteins and SR-related proteins participate in constitutive and alternative splicing of pre-mRNA, generally as enhancers of the splicing reaction (Blencowe, 2000), but in some cases they can also enhance 3'end processing (Lou et al., 1998; Lou et al., 1995; Maciolek and McNally, 2007; McCracken et al., 2003). Interestingly, the purine-rich sequence present in the enhancer element was very similar to that one recognized by the SR protein SRSF1 (Caputi et al., 1994; Lavigueur et al., 1993). As mentioned above, the insertion of the 166 bases-long enhancer element was able to rescue the signal on the Northern blot, but not to the full activity of the complete **A1** region (Figure 30B and D, Lane 4 compared with Lane 1). Therefore, I wondered whether SR proteins could play a role in 3'end processing at the PAS4. Overexpression of SR proteins may increase their binding to the purine-rich region of the enhancer element affecting the 3'end processing efficiency at the PAS4, which can either cause an increase or decrease in the amount of the mRNA. To this end, HeLa cells were transfected with the **A1[**215-380**]-A4** construct together with a plasmid expressing the SRSF1 splicing factor. The northern blot analysis showed a decrease in the amount of mRNA when SRSF1 was overexpressed (Figure 32B, C and D). Therefore, SRSF1 may have a negative role in 3'end processing at the PAS4 or, alternatively, the enhanced binding of the SRSF1 protein to the purine rich-region may compete with the binding of some other factor necessary to confer full activity to the enhancer element. Further experiments have to be performed to determine the precise role of SRSF1 in beta adducin polyadenylation.

In order to identify and characterize the trans-acting factor/s recognizing the enhancer

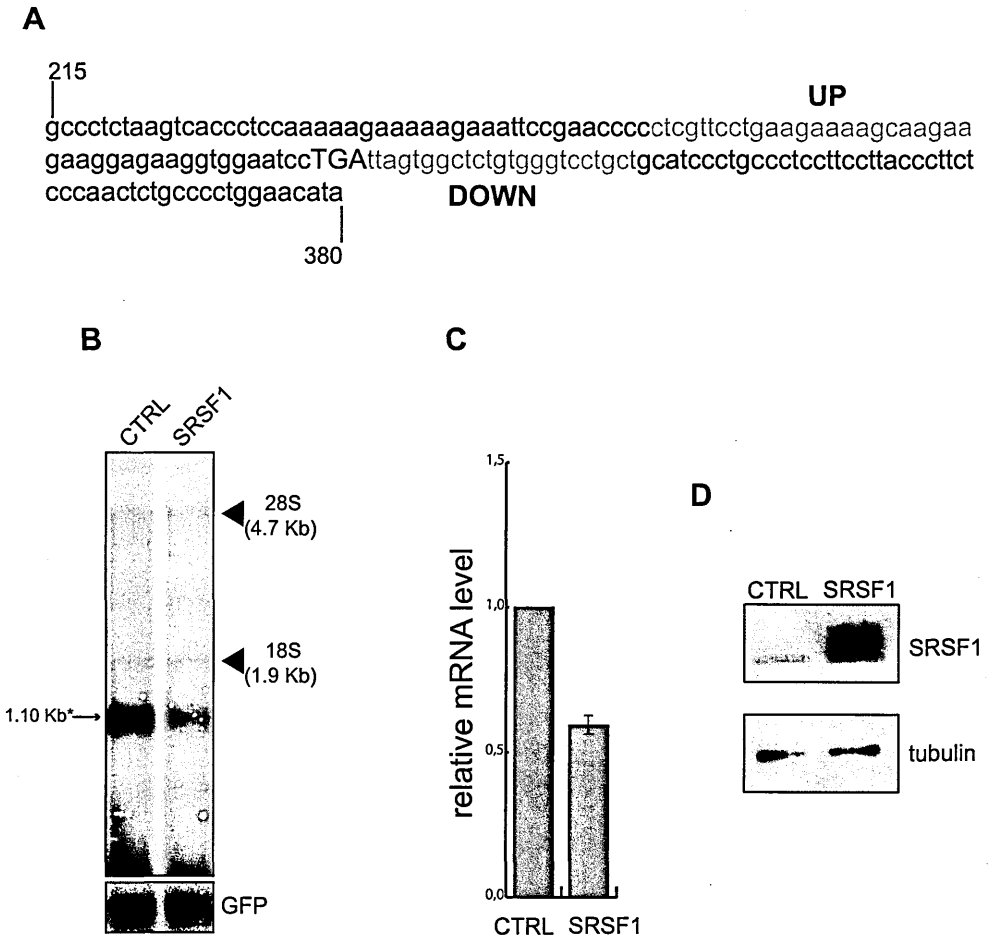


Figure 32
Effect of SRSF1 overexpression on minigene expression.
A: Sequence composition of the 166 nt long enhancer element; the STOP codon (TGA) is indicated in capital letters ; UP and DOWN are upstream and downstream elements marked by orange letters; numbers correspond to the nucleotide position respective to the first base of the last mouse beta adducin exon. **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the **A1**[215-380]**-A4** constructs together with pCG plasmid (CTRL) or pCG-SF2/ASF plasmid containing SRSF1 cDNA (SRSF1); the alpha globin probe was used to detect the mRNAs of interest; the p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected mRNA size is indicated on the left; (i.e., the asterisk indicates without the pA tail); **C**: The relative mRNA levels calculated as the ratio between analyzed mRNAs and GFP control mRNAs shown in Panel B; quantification of the Northern blot signals has been performed using OptiQuant software; the mean +/- S.D. of three independent experiments is shown; **D**: Western blot analysis for SRSF1 of the protein extracts prepared from HeLa cells previously transfected as described in Panel B; tubulin was used as a loading control.

elements, I have performed some preliminary analysis of the proteins interacting with the 69 bases-long enhancer element (Figure 29A). These studies were done in parallel with the analysis of the functionality of this element. In a first step to identify the *trans*-acting factors recognizing the 69 bases-long enhancer element, I performed Electrophoretic Mobility Shift Assays (EMSA) and RNA-Pull-down experiments. The RNAs used for the analysis were generated by *in vitro* transcription of DNA templates. I used three different DNA templates to generate the radioactively labeled RNAs: a) a 69 bases-long region

that contained the complete enhancer sequence, with the upstream and downstream enhancer elements (***WT***); b) a 45 bases-long region that lacked the upstream element (***Δup***); c) a 46 bases-long region that lacked the downstream element (***Δdown***) (Figure 33A). These *in vitro* transcribed RNAs formed secondary structures which were efficiently digested with RNase V1, an enzyme that specifically degrades double stranded RNA (Figure 33B, first three lanes).

I have determined the ability of these RNA sequences to bind a protein or protein complexes by the EMSA assay. The radioactively labeled RNAs were incubated with HeLa nuclear extract. RNA-protein interactions were then detected as a band shift on a native polyacrylamide gel electrophoresis. All three tested RNAs (***Δup*** ,***WT*** and ***Δdown***) were able to bind proteins producing lower mobility bands (Figure 33B). Two RNA-protein complexes were observed for the ***Δup*** probe, while only one complex was observed with the ***WT*** and ***Δdown*** probes. To determine the specificity of the complexes and to test whether these RNAs were bound by the same protein complexes, I have performed EMSA competition assays (Figure 34). The labeled RNA probe was incubated with HeLa nuclear extract in the presence of a non-radioactive competitor (cold RNA). As it can be seen in Figure 34, all cold competitors were able to compete out the complexes formed by their corresponding hot RNA probes, confirming the specificity of the complexes. The cold ***WT*** competitor was also able to compete out the complexes formed with the hot ***Δdown*** RNA, as expected, but not those formed with the hot ***Δup*** probe (Figure 34C and 34B, respectively). In addition, the cold ***Δup*** competitor was not able to compete for the proteins bound to the hot ***WT*** probe. The cold ***Δdown*** competitor was able to compete for the proteins bound to the hot ***WT*** probe (Figure 34A). To conclude, it is likely that both the ***WT*** and ***Δdown*** probes were able to form similar RNA-protein complexes, which differ from the ones formed by the ***Δup*** probe. These differences could be directly related to the folding structure of each RNA and/or to the impossibility of crucial proteins to take part in the formation of the complex due to the deletion of upstream enhancer element.

A

UP DOWN

gaaggagaaggtggaatccTGAttagtggctctgtgggtcctgct Δ*up* 45nt

ctcgttcctgaagaaaagcaagaagaaggagaaggtggaatccTGAttagtggctctgtgggtcctgct *wt* 69nt

ctcgttcctgaagaaaagcaagaagaaggagaaggtggaatccTGA Δ*down* 46nt

B

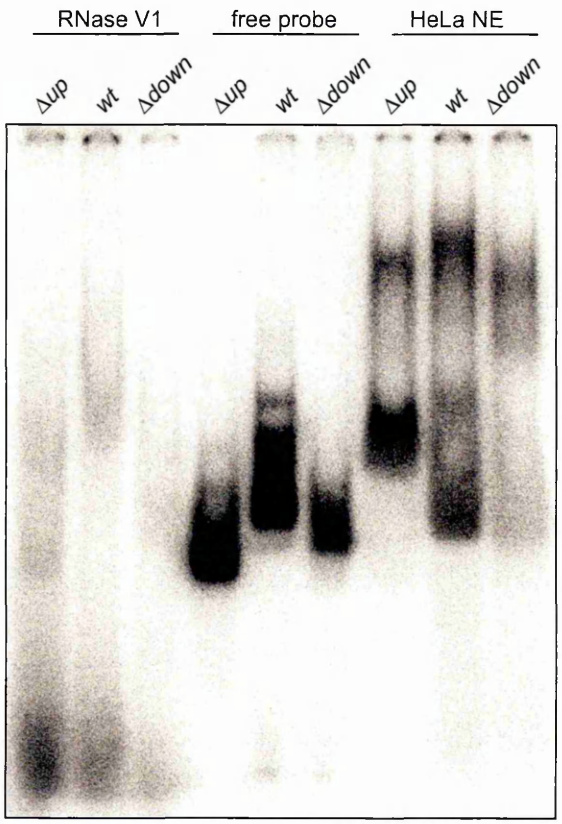
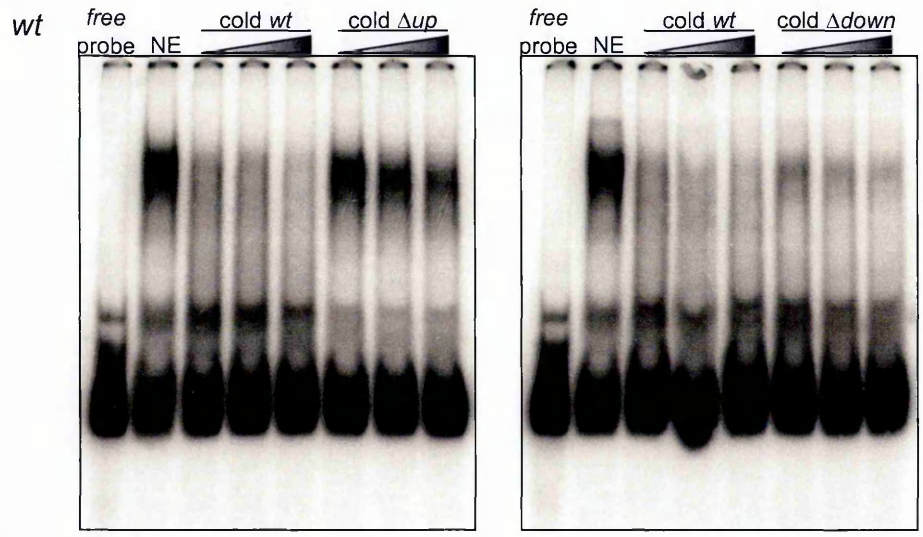


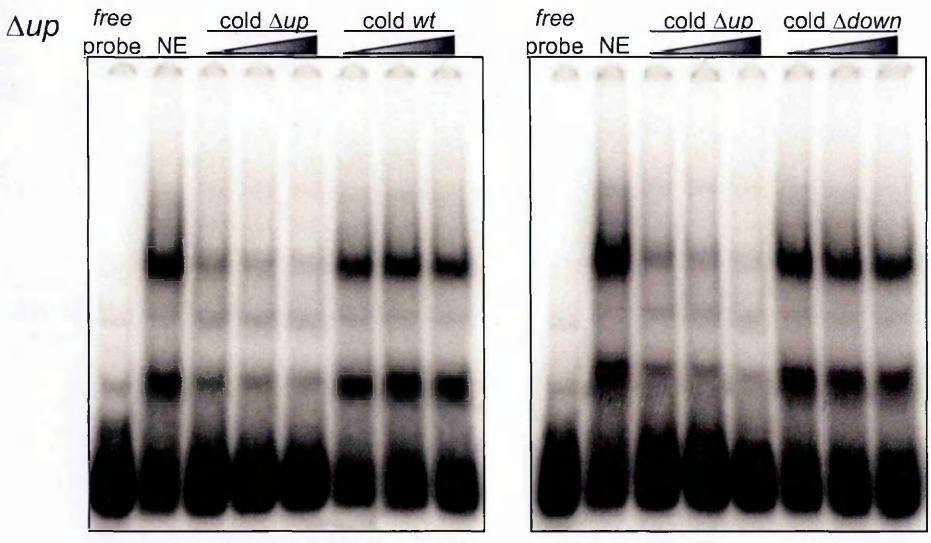
Figure 33
RNA EMSA analysis.
A: DNA sequences used for the in vitro transcription reactions to produce corresponding RNAs (Δ***up***, ***wt*** and Δ***down***). **B**: RNA-protein interaction analysis, EMSA assay; the first three lanes represent digestion of the secondary structured hot RNAs (Δ***up***, ***wt*** and Δ***down***) by double-stranded RNA specific ribonuclease, RNAse V1; the second three lanes represent the hot RNAs; last three lanes represent interaction between the hot RNAs and the proteins from the HeLa nuclear extract.

In parallel, to identify the proteins bound to the enhancer element, I have performed preliminary RNA-Pull-down experiments. The RNA probe was bound to agarose beads, followed by the binding reaction using HeLa nuclear protein extracts. Protein complexes bound to a given RNA sequence were separated from the unbound proteins present in the binding reaction, eluted, separated on SDS-polyacrylamide gels and finally stained with colloidal coomassie blue. Then, differences in the protein pattern could be rela-

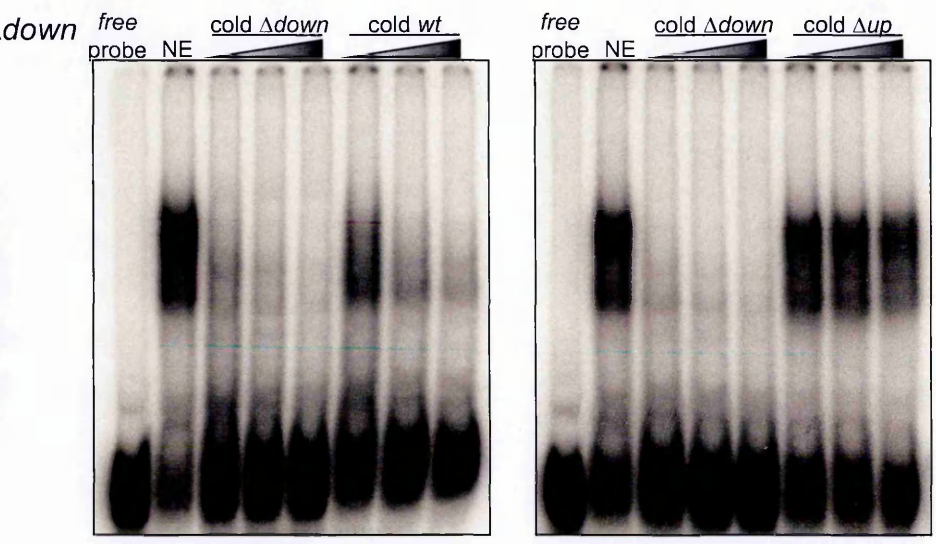
A

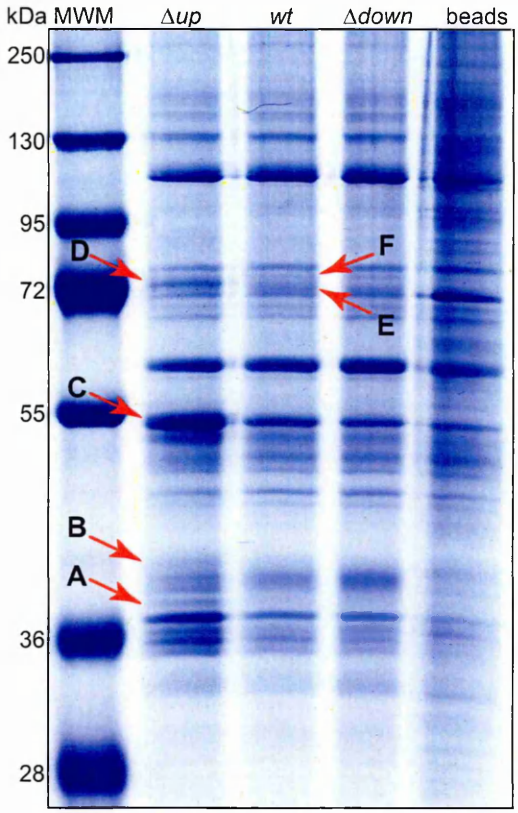

wt
free probe
NE
cold wt
cold Δup
free probe
NE
cold wt
cold Δdown

B


Δup
free probe
NE
cold Δup
cold wt
free probe
NE
cold Δup
cold Δdown

C


Δdown
free probe
NE
cold Δdown
cold wt
free probe
NE
cold Δdown
cold Δup



A	- hnRNP A2B1 (score 344) - hnRNP H3 (score 315) - hnRNP A3 (score 168)
B	- hnRNP A3 (score 235) - hnRNP A2B1 (score 94) - hnRNP A1 (score 94)
C	- hnRNP H1 (score 890) - hnRNP H2 (score 473) - hnRNP F (score 206)
D	- FUBP1 (score 822) - hnRNP M (score 269)
E	- hnRNP M (score 516) - FUBP1 (score 69)
F	- RBM 14 (score 186) - hnRNP M (score 166)

score >65 significant, p<0.05

Figure 35
Pull-down analysis of protein bound to the enhancer element.
Left panel - RNA-Pull-down assay for the Δ***up***, ***wt*** and Δ***down***; DNA sequences used for the in vitro transcription reactions to produce corresponding RNAs (Δ***up***, ***wt*** and Δ***down***) are as described in Figure 33; in vitro synthetized RNAs were covalently linked to adipic-acid beads and subsequently incubated with HeLa nuclear extract; the unbound proteins were washed and the ones that remained bound to the RNAs were eluted, separated on 10% SDS-PAGE and visualized by colloidal coomassie blue staining; bands representing single protein differences are indicated with red arrowheads and by letters (A, B, C, D, E and F); MWM – molecular weight marker. Right panel - each of the bands represented in the left panel (A-F) was cutted and the corresponding protein identified by mass spectometry (A-F).

<< Figure 34
RNA EMSA competition analysis.
A: competition of the cold Δ***up***, ***wt*** and Δ***down*** RNAs with the hot ***wt***. **B**: competition of the cold Δ***up***, ***wt*** and Δ***down*** RNAs with the hot Δ***up***.**C**: competition of the cold Δ***up***, ***wt*** and Δ***down*** RNAs with the hot Δ***down***. 0.5, 1 and 2 pmol of cold competitor were used.

ted to differences in the RNA-protein complex formation for each RNA probe used. In line with the EMSA analysis, protein profiles of the ***WT*** and ***Δdown*** were similar. Single protein differences were observed between the pull down profiles obtained with the ***Δup*** and ***WT*** RNAs (Figure 35, the red arrows point to the differences), which were then analyzed by mass spectrometry (Figure 35, right panel). Most of the identified proteins belonged to the hnRNP family of proteins. These proteins are trans-acting factors that bind pre-mRNA, modulate pre-mRNA splicing and in some cases also polyadenylation, such as hnRNP H (Arhin et al., 2002; Wilusz and Beemon, 2006).

2.3 Study of the functions of the spleen- and brain-specific 3'UTRs of the beta-adducin mRNA

To analyze the potential effects of the spleen- and brain-specific 3'UTRs on the beta-adducin mRNA levels I have introduced the different 3'UTRs of the mouse beta-adducin gene (ending at the PAS1, PAS23 and PAS4 cleavage sites) into the pGL3-control reporter vector, downstream of the firefly luciferase gene (Figure 36A). Downstream of the inserted 3'UTRs the SV40 late polyadenylation signal of the reporter vector was present, which assured efficient cleavage and polyadenylation of the reporter gene. Epithelial (HeLa - human cervical carcinoma), neuronal (SK-N-BE - human neuroblastoma) and hematopoietic (MEL-murine erythroleukemia) cells were transfected with these constructs together with the pHRG-TK vector that contains the renilla luciferase (RL) gene. The RL construct was used to normalize for differences in transfection efficiency. Luciferase activity was tested by performing the Dual-Luciferase Reporter assay and the relative luciferase activity was used as a measure of expression efficiency (Figure 36B). For all the different cells types tested, the same pattern of luciferase expression was observed: a decrease in the relative luciferase activity correlated with increasing length of the 3'UTR.

In order to investigate whether the observed differences in luciferase activity (Figure 36B) were due to changes in translational efficiency or stability of the mRNAs, I performed RT-qPCR. The results showed a decrease in the mRNA levels in the constructs having longer 3' UTRs, suggesting that the main mechanism acting on the regulation of luciferase expression was through mRNA stability (Figure 36C). However, the decrease in mRNA levels did not account for all the differences in luciferase observed between the different constructs

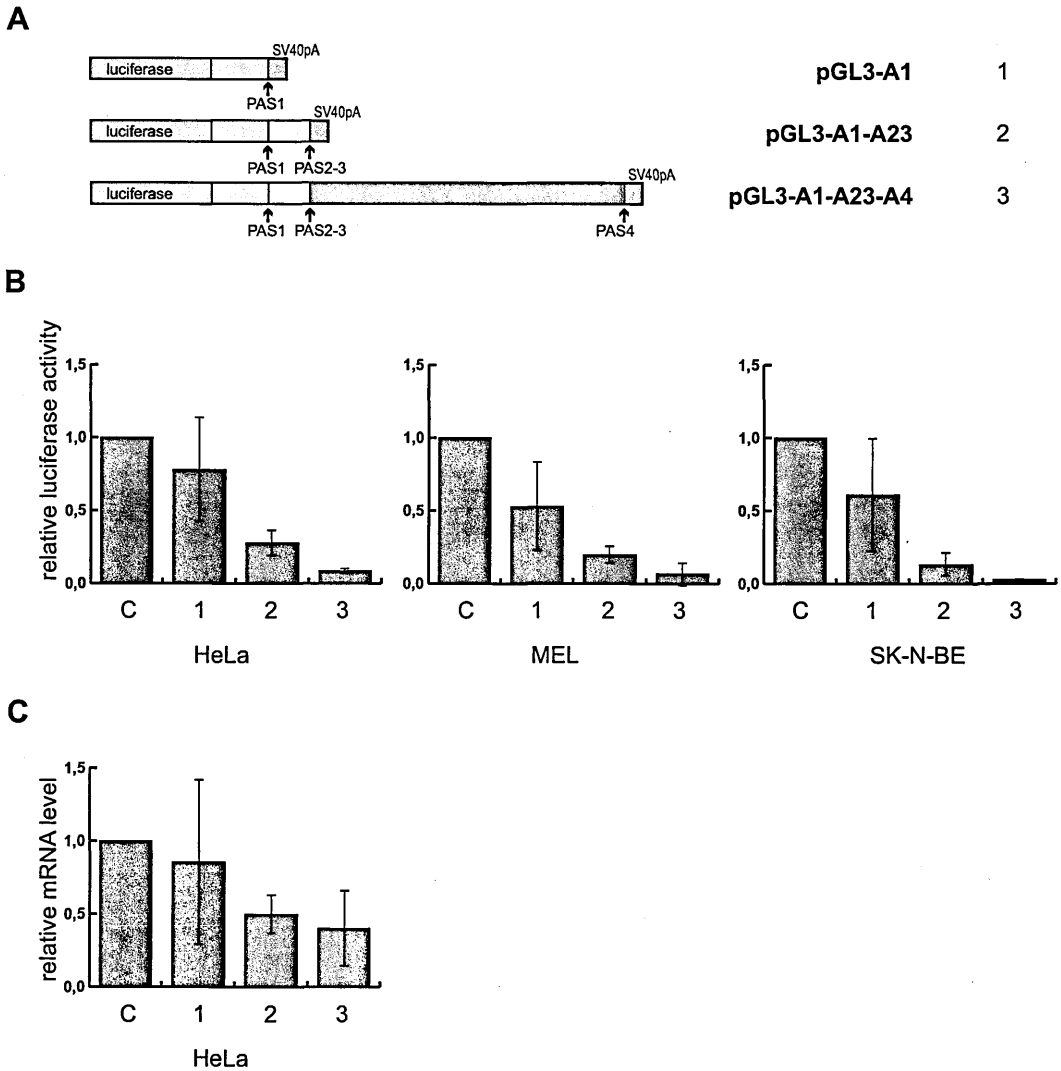


Figure 36

Reporter gene analysis showed that expression of the long beta adducin 3'UTR is repressed.

A: Schematic representation of the pGL3 constructs: pGL3-A1 (1), pGL3-A1-A23 (2) and pGL3-A1-A23-A4 (3); full length of different 3'UTRs of mouse beta adducin are placed downstream of the firefly luciferase gene of pGL3-control vector; different 3'UTR fragments containing the tissue-specific polyadenylation sites of mouse beta adducin (PAS1, PAS23 and PAS4) are indicated as orange, yellow and blue boxes, respectively (see matherials and methods for details; the enhancer sequence element containing the stop codon from beta-adducin that is defined in the above mentioned studies is not included in the construct); the green boxes indicate luciferase gene; the grey boxes indicate SV40 late polyadenylation signal (SV40pA); **B**: Graphic representation of the relative luciferase activity in the protein extracts of HeLa, MEL and SK-N-BE cells previously transfected with the pGL3 empty control (C) and constructs shown in panel A, together with pHRG-TK vector that contains the renilla luciferase gene; the mean +/- S.D. of nine, four and five independent experiments for HeLa, MEL and SK-N-BE, respectively, is shown; experiments were performed in triplicates. **C**: RT-qPCR analysis of the relative mRNA levels calculated as the ratio between the amounts of the firefly and renilla luciferase mRNAs from total RNA prepared from HeLa cells previously transfected with the pGL3 empty control (C) and constructs shown in panel A, together with pHRG-TK vector; the mean +/- S.D. of three independent experiments is shown; expreriments were performed in duplicates.

(Figure 36B), suggesting that other mechanisms regulating mRNA translation may be also involved. In addition, we should also consider that the efficiency of Pol II in transcribing the longer templates may be lower, which may result in lower levels of mRNAs of the constructs having the longer 3' UTRs.

As already mentioned in the Introduction section, the region containing the PAS4 of the beta-adducin is highly conserved between mammals (Figure 14 and 15). The region of conservation contains, in addition to the core polydenylation elements, other potential regulatory elements, such as CPE and ARE. There are two potential CPEs aproximately 200 bases upstream of the hexanucleotide motif of the PAS4. Recently has been demonstrated that the CPE-containing mRNAs can be targeted by CPEB also in the nucleus (Lin et al., 2010). The nuclear function of CPEB is still not elucidated, although there are some suggestions for its implication in alternative splicing (Lin et al., 2010) and polyadenylation (Raul Mendez, personal communication). The CPE role in alternative polyadenylation has been proposed for the BUB3 gene where two CPEs close to the proximal PASs were involved in the switching to a more distal PAS (Raul Mendez, personal communication). In the context of beta-adducin, although no CPE consensus sequences were present in the 3' UTR proximal region, I tested whether the mutation of the two distal CPEs could trigger the usage of the proximal PASs. To this end, the CPE mutations were introduced into the **A1-A23-A4** construct to generate the **A1-A23-A4[**mutCPE**]** construct (Figure 37A). The expression profile of this construct was analyzed by Northern blot of total RNA prepared from transfected HeLa cells. In the case the CPEs were regulating alternative polyadenylation of beta-adducin, a new band corresponding to polyadenylation in the proximal PAS should appear in the Northern blot. However, as shown in Figure 37B, only one signal was obtained which corresponded to the usage of the PAS4. Therefore, we can conclude that CPEs were not regulating beta-adducin alternative polyadenylation. On the other hand, there was a twofold increase in the levels of mRNA generated by the construct containing the mutated CPEs (Figure 37C). This result suggests a role of CPEs in mRNA stability of beta-adducin.

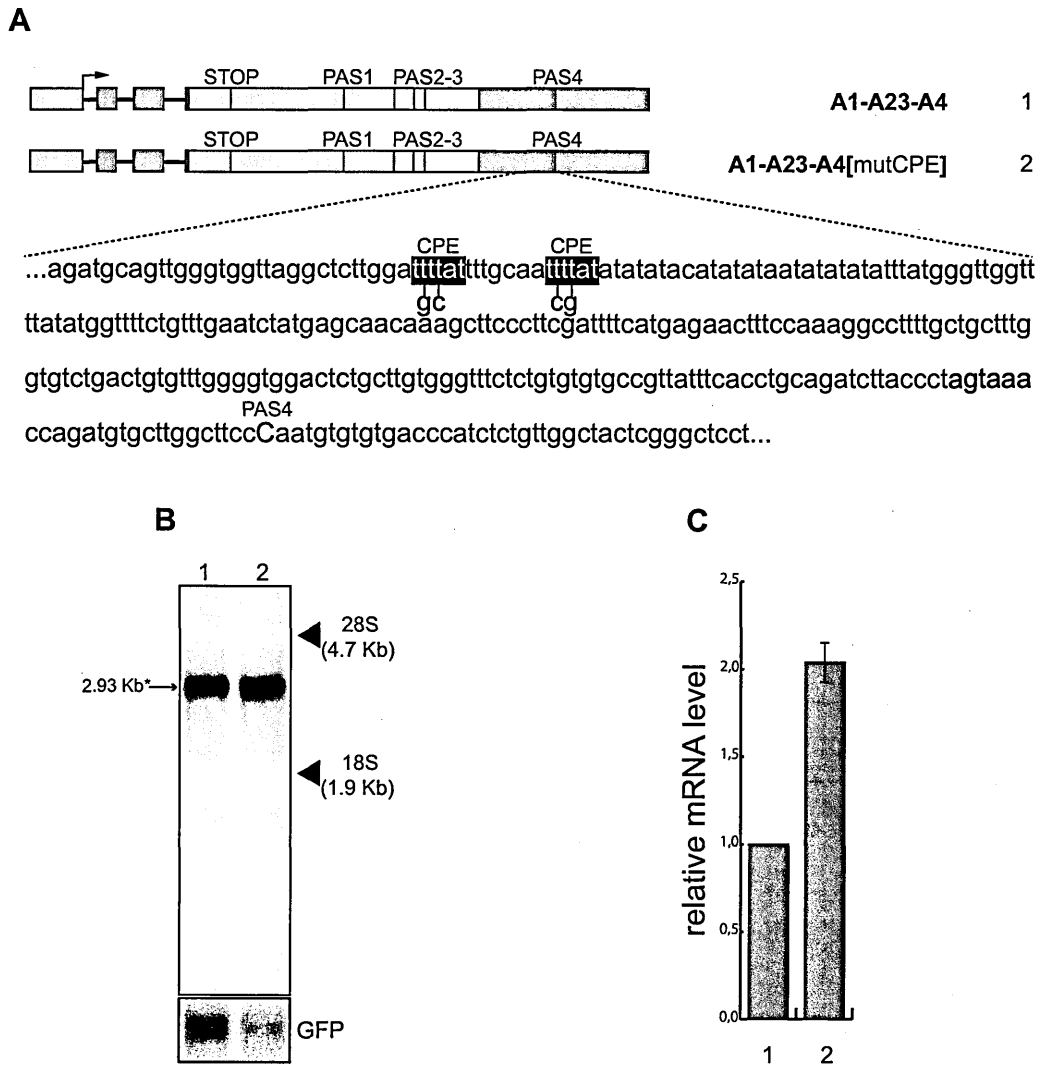


Figure 37

Mutation of CPEs results in the increase of relative mRNA levels.

A: Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A1-A23-A4**[mutCPE] (2); the sequence containing CPEs is shown below; the black boxes indicate CPEs; the mutations introduced inside the CPEs are indicated in red lowercase letters; the hexanucleotide motif sequence and the cleavage site (PAS4) are indicated in blue and red lowercase letters, respectively. **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in Panel A; the alpha globin probe was used to detect the mRNAs of interest; p-EGFP-C2 plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected mRNA size is indicated on the left; (i.e., the asterisk indicates without the pA tail); **C**: The relative mRNA levels calculated as the ratio between analyzed mRNAs and GFP control mRNAs shown in Panel B; quantification of the Northern blot signals has been performed using the OptiQuant software; the mean +/- S.D. of two independent experiments is shown.

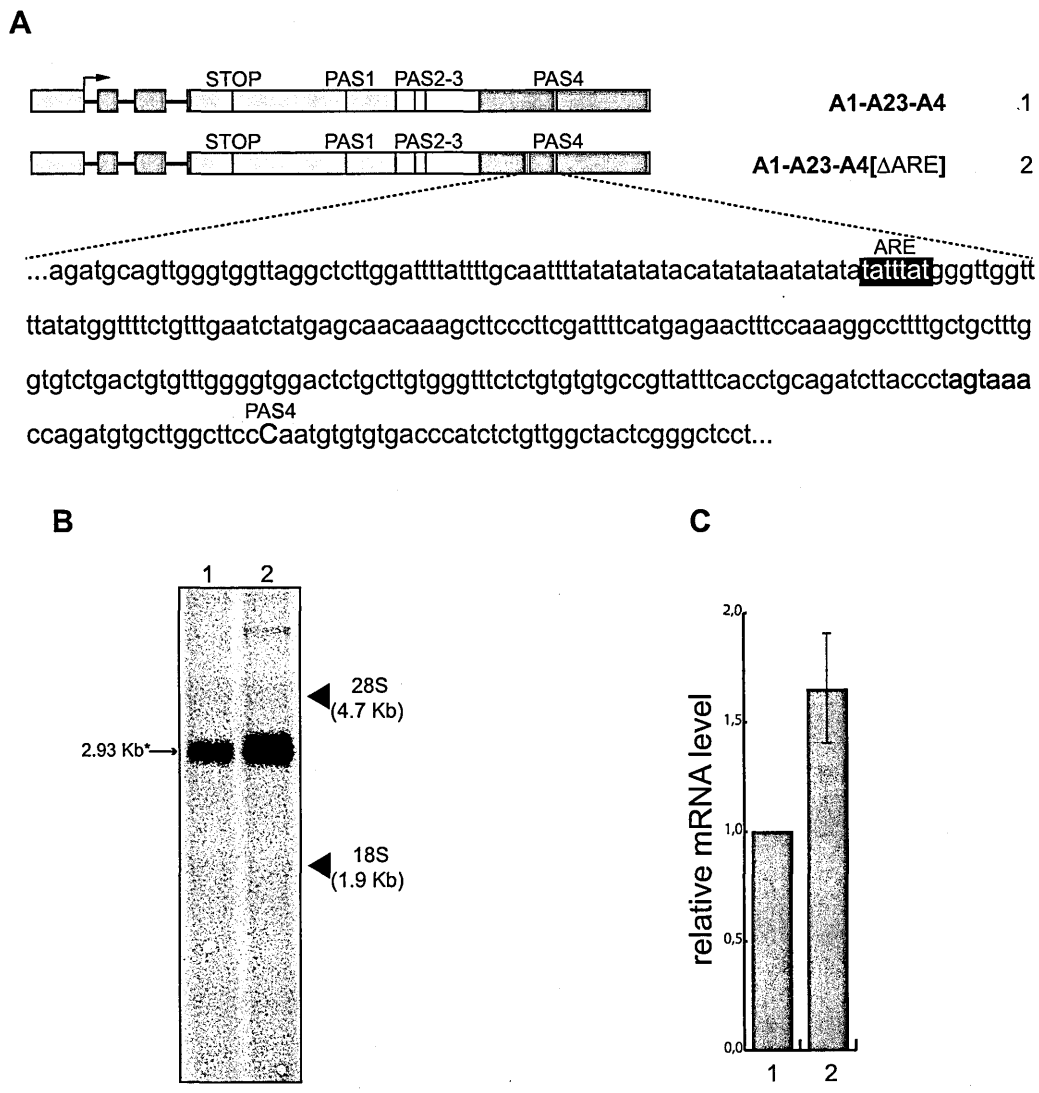


Figure 38

Deletion of ARE results in the increase of relative mRNA levels.

A: Schematic representation of the chimeric minigene constructs: **A1-A23-A4** (1), **A1-A23-A4[**ΔARE**]** (2); the sequence containing ARE is shown below; the black box indicates the ARE, which has been deleted in the **A1-A23-A4[**ΔARE**]** construct; the hexanucleotide motif sequence and the cleavage site (PAS4) are indicated in blue and red lowercase letters, respectively; the white box in the construct indicates deleted ARE. **B**: Northern blot analysis of total RNA prepared from HeLa cells previously transfected with the constructs shown in Panel A; phRG-TK (Renilla) plasmid was used for the normalization of the transfection efficiency; the positions of the 28S and 18S rRNAs are indicated on the right; the expected mRNA size is indicated on the left; (i.e., the asterisk indicates without the pA tail); **C**: RT-qPCR analysis of the relative mRNA levels calculated as the ratio between analyzed mRNAs and Renilla control mRNA; the mean +/- S.D. of three independent experiments is shown; experiments were performed in duplicates.

We observed that the region upstream of the PAS4 is an AU-rich region, where a potential ARE was observed (Figure 38A). To test the functionality of this potential ARE, this element was deleted from the **A1-A23-A4** construct to generate the **A1-A23-A4[**ΔARE**]** construct (Figure 38A). The expression of the deleted construct was tested by transfection into HeLa cells. The RL plasmid was cotransfected and its expression levels determined by RT-qPCR to normalize for differences in the transfection efficiency.

The Northern blot analysis showed normal processing at the PAS4 (Figure 38B, Lane 2). Deletion of the potential ARE resulted in ~60% increase in the relative amounts of mRNA levels (Figure 38C), suggesting a role of the deleted ARE in destabilization of the beta-adducin mRNA.

To conclude, these results suggest that the RNA levels of the long beta adducin transcripts are regulated, at least partially, by modulating the stability of the mRNA through CPE and ARE, elements present only in the mRNA isoform which is polyadenylated in the distal PAS. In fact, mutation of CPE and deletion of ARE resulted in a significant increase in mRNA levels of the chimeric minigene constructs. Other mechanisms, such as regulation by miRNAs, could also account for the regulation of beta-adducin gene expression.

Chapter 3

Discussion and conclusions

The presented results underscore the complexity of the mechanisms that regulate polyadenylation of the mouse beta adducin pre-mRNA. In order to study the mechanisms acting on the beta adducin distal PAS4, we first identified the core cis-acting elements (Hm, DSE and USE) which were necessary to confirm usage of the PAS4. Surprisingly, I found that additional long-distance enhancer and silencer cis-acting elements are involved in the regulation of 3'end processing at the PAS4. To the best of our knowledge, this is the first demonstration of the presence of long-distance upstream polyadenylation enhancer and silencer elements in eukaryotic, non-viral, pre-mRNAs.

3.1 Cis-acting elements regulating the mouse beta adducin PAS4 usage

3.1.1 Mouse beta adducin PAS4 is preferentially used in HeLa cells

In an array of tandem PASs, the distal PAS tends to be the strongest and most efficiently used PAS (Beaudoing et al., 2000; Legendre and Gautheret, 2003). This tendency was also observed in our minigene model containing all three beta adducin polyadenylation sites. When transfected into HeLa cells, the distal PAS4 was exclusively, similar to that observed in brain. The preferentially usage of the distal PASs in respect to proximal ones in HeLa cells is detected for several endogenous genes (Kubo et al., 2006), suggesting that the usage of distal PASs is generally preferred in this cell line. The observed inefficient 3'end processing at the beta adducin proximal PAS1 and PAS2-3 could be explained by the presence of low-conserved and poorly defined cis-acting elements regulating their usage. On the contrary, *in silico* analysis of the PAS4 region showed the presence of highly conserved and well-defined polyadenylation cis-acting elements, whose role was confirmed in the presented study. Additionally, the recognition and usage of the PAS1 and PAS2-3 might depend on the presence of tissue-specific trans-acting factors which may be absent in HeLa cells. Thus, the usage of these tissue-specific sites could be expected to occur in a more physiological system. Unfortunately, technical problems prevented us from doing so, as described above in the section Results.

3.1.2 Canonical cis-acting elements regulate mouse beta adducin PAS4 usage

Deletion analyses demonstrated that the *in silico* predicted sites were indeed the correct functional polyadenylation core elements defining PAS4 (Figure 17). In accordance,

absence of either the Hm or the DSE abolished 3'end processing at the PAS4 of the beta adducin pre-mRNA. However, inactivation of the PAS4 by either deleting the Hm or the DSE, has not led to the activation of the proximal PAS1 and/or PAS2-3. Instead, two cryptic PASs located about 200 bases upstream of the PAS4 have been used. In addition, the proximal PASs were inactive even in the absence of the entire **A4** region (data not shown). Therefore, the absence of polyadenylation activity at the PAS1 and PAS2-3 was not due to the presence of the stronger PAS4 in the same transcript. Most likely additional trans-acting factors, which may be absent in HeLa cells, are necessary for the efficient 3'end processing at the proximal PAS1 and PAS2-3 of beta adducin pre-mRNA. The identified Hm, located 18 nucleotides upstream of the PAS4, is represented by the non-canonical and less frequent **AGUAAA** sequence. Immediately downstream of the PAS4, a 29 bases long **GU**-rich region represents the identified DSE. Recognition of the Hm and DSE by CPSF and CstF, respectively, represents the main characteristic in the initiation of 3'end processing factors assembly on the polyadenylation signal. Purified CPSF is known to bind the Hm weakly and this binding is enhanced by its interaction with the CstF bound to the DSE (Gilmartin and Nevins, 1991; MacDonald et al., 1994; Weiss et al., 1991). Here, we observed the dependence of the Hm on the presence of the DSE and vice versa, in the recognition of the PAS4. As mentioned above, absence of one of these polyadenylation elements disrupted the efficient recognition of the PAS4 and led to the usage of two upstream cryptic PASs. These two cryptic PASs are located in a highly conserved region at the base positions 5742 and 5793, in respect to the first base of the mouse beta adducin terminal exon. Upstream of both cryptic PASs, we identified the noncanonical **AAUAUA** hexamer, which together with various **U** and **GU**-rich stretches located downstream, may play a role in the recognition of both cryptic PASs. The **AAUAUA** hexamer is detected in ~2% of human and mouse PASs, and is ~ 1.5 folds less active than the **AGUAAA** hexamer and 10 fold less active than the canonical **AAUAAA** hexamer in *in vitro* polyadenylation assays (Sheets et al., 1990; Tian et al., 2005)(Figure 3). Efficiently processed PASs are usually associated with

a high number of polyadenylated ESTs ending at the PAS position, and those PASs are usually the ones bearing canonical hexamer variants, **AAUAAA** and **AUUAAA** (Beaudoing et al., 2000). Interestingly, approximately the same mean number of ESTs was found to be associated with PASs containing the canonical **AUUAAA** and noncanonical **AGUAAA** hexamer (Beaudoing et al., 2000), suggesting that the **AGUAAA** hexamer can be as active as the canonical **AUUAAA**. In the case of beta adducin, a higher number of ESTs was observed to align with the PAS4 (containing the **AGUAAA** hexamer) than with the PAS1 (containing the **AUUAAA** hexamer), suggesting that the processing at the PAS4 is more efficient (Costessi et al., 2006). However, a higher number of ESTs does not necessarily corelate with PAS efficiency. A number of ESTs associated with a PAS also depends on the number of analysis performed for each tissue that are available in the databases, if the sites that one want to compare are tissue-specific. Additionally, mRNAs processed at the cryptic PASs were expressed at lower levels than the ones processed at the canonical PAS4, suggesting a less efficient processing at these cryptic PASs.

Interestingly, regulatory elements such as ARE and CPE were not lost in the 3'UTR generated from the usage of the cryptic PASs. Therefore, the close proximity of the cryptic PASs favours the hypothesis of a mechanism for the generation of a long 3'UTR even in case of the failure of the 3'end processing at the strong PAS4. Similar 3'end processing fail-safe mechanisms for the eukaryotic pre-mRNAs have not been described previously. It would be of interest to extend the studies to other genes whose polyadenylation signal sequences composition show similarities to that of beta adducin.

In addition to the importance of the Hm and DSE, we demonstrated that the presence of the USE was also important for the efficient processing at the PAS4. The general role of the USEs is in the enhancement of 3'end processing. Here, we showed that a 32 bases long **U/GU**-rich region located upstream of the Hm represents a USE acting on the PAS4, as the deletion of this element led to the less efficient processing at PAS4 manifested by a decrease in mRNA levels (Figure 17B).

3.1.3 Novel non-canonical long-distance upstream enhancer and silencer elements regulate mouse beta adducin PAS4 usage

Although if the core cis-acting elements we described above were critical for processing at the PAS4, they were not sufficient to trigger processing at PAS4. In fact, the presence of the upstream 1.4 kb long **A1** region appeared to be indispensable for 3'end processing at the PAS4, as in its absence no processing at the PAS4 was observed (Figure 21). Mapping experiments showed that the minimal enhancer sequence able to activate the usage of the PAS4 with the same efficiency as the full lenght 1.4 Kb **A1** region was contained within a 355 nt sequence (Figure 28, construct **A1[**17-371**]-A4**). These 355 nt comprise of the last part of the coding region, the stop codon and the first 66 nt of the 3'UTR, spanning from base 17 to base 371 of the mouse beta adducin terminal exon. Shorter fragments of 166 nt and 69 nt have been tested, but they were less efficient in activating the PAS4. The deletion experiments showed that this non-canonical enhancer element is a complex element, composed of various different sub-elements, each of which contribute to the overall function of the enhancer resulting in the activation of the PAS4. Several characteristics distinguish this polyadenylation upstream enhancer element from the more "classical" USEs that were previously described.

Firstly, the beta adducin enhancer element is located more than 2 Kb upstream of the PAS4 in the minigene context (construct **A1-A23-A4**). In the natural mouse beta adducin context the enhancer is positioned more than 5 Kb upstream of the PAS4. Although we have not yet proven the functionality of the enhancer element in its natural context, we expect it will be active because the enhancer activity was not affected by the distance between the enhancer and the PAS4, as verified by several tested constructs (Figures 27 and 28). The presence of the long-distance upstream polyadenylation enhancer element is quite peculiar for eukaryotic polyadenylation signals. In general, the USEs are positioned immediately upstream of the Hm. Interestingly, a long-distance upstream polyadenylation enhancer element has been observed for the Rous sarcoma virus (RSV) polyadenylation signal (Fogel et al., 2002; Wilusz and Beemon, 2006). This element is called negative re-

gulator of splicing (NRS) and is located in the gag gene, about 8 Kb upstream of the RSV polyadenylation signal. Binding of the SR proteins SRSF1, SRSF3 and SRSF7 to the NRS stimulates polyadenylation at the RSV PAS (Maciolek and McNally, 2007). The SR proteins are in competition for the binding to the NRS with hnRNP H, which inhibits polyadenylation at the RSV PAS (Wilusz and Beemon, 2006). Whether the presence of the long-distance USE represents an exception or a general property of eukaryotic PASs is yet to be studied. Secondly, the non-canonical enhancer element is also upstream of the alternative polyadenylation sites but only influences the usage of PAS4. The activity of the PAS1 and PAS2-3 are not affected by the enhancer even in the absence of the **A4** region. Thus, it is unlikely that this element acts as a general enhancer of polyadenylation and may be specific for the beta adducin gene.

Thirdly, the non-canonical enhancer element has a complex structure spanning immediately upstream and downstream of the stop codon of the mouse beta adducin terminal exon. Thus, this element partially spans throughout the coding region, a USE characteristic that has not previously been reported. Being positioned in the coding region, the sequence composition is subject to evolutionary constrains dictated by the protein it encodes which makes it unlikely to be a general enhancer element. Usually, the USEs are represented by **U**-rich sequences, while the coding region of the non-canonical enhancer element is purine-rich and contains a high number of potential binding sites for SR-proteins. Overexpression of SRSF1, an SR-protein known to be involved in stimulating 3'end processing, resulted in no enhancement of beta adducin PAS4 pre-mRNA processing. On the contrary, a twofold decrease in the mRNA levels of the construct containing the 166 nt long enhancer element has been detected , suggesting a decrease in the efficiency of processing at the PAS4 (Figure 32). A possible explanation of this observation could be that the overrepresented SRSF1 in the transfected cells competed out the binding of other factor(s) to the enhancer element, necessary for the activation of the PAS4 processing. However, further experiments are needed to confirm this possibility. One of the potential protein candidates could be the splicing factor SRSF10 (also

known as Tra2 beta), as its consensus binding sequence is similar to that of the SRSF1 (Grellscheid et al., 2011; Sanford et al., 2009). Manual inspection of the 355 nt enhancer element sequence revealed the presence of 13 out of the 20 most frequent hexamers found in CLIP-tags of SRSF10 (Figure 39)(Grellscheid et al., 2011). There is no evidence for a potential role of SRSF10 in polyadenylation, apart from its Drosophila homolog, which is able to enhance female specific splicing and polyadenylation in the dsx pre-mRNA (Hedley and Maniatis, 1991). Nevertheless, the interaction between SRSF10 and the 3'end processing factor CFI-68 has been detected in HeLa and HEK293 cell lines (Dettwiler et al., 2004). Therefore, SRSF10 represents a good candidate for binding to the non-canonical enhancer element. It would be of interest to test its role in 3'end processing of the mouse beta adducin PAS4. However, it should be taken into acount that none of the pulled-down proteins analyzed was an SR protein (Figure 35). Improving the experimental conditions could reveal more protein candidates for the binding to the enhancer element. On the contrary, hnRNP proteins were detected in the bands isolated from the pull down experiments, among which I found hnRNP H, already known to be involved in 3'end processing. Binding of hnRNP H protein to **G**-rich auxiliary DSEs stimulates cleavage and polyadenylation (Arhin et al., 2002; Bagga et al., 1998; Dalziel et al., 2007). A recent study has shown that hnRNP H bound to **G**-rich elements located upstream of the PAS stimulates polyadenylation through the interaction with poly(A) polymerase (Millevoi et al., 2009). Whether hnRNP H has a role in 3'end processing at the beta adducin PAS4 will be addressed in future studies.

I have shown that the PAS4 requires the presence of an upstream non-canonical enhancer element. The effect of the enhancer element was lost when located downstream of the "proximal" PAS4, as observed in the **A4-A1-A23-A4** construct (Figure 18). A position-dependent effect of the polyadenylation enhancer element has also been observed for beta globin (Antoniou et al 1998). The last 60 nt of the second beta globin intron represent the enhancer element and it was not functional when placed downstream of the beta globin PAS. However, the beta globin enhancer element is part of the

A

6-mer	Frequency of 6-mer in CLIP tags (S)	Frequency of 6-mer in genome	Genome corrected frequency (G)	Ranking (G)	Frequency of 6-mer in mouse testis transcriptome	Transcriptome corrected frequency (T)	Ranking (T)
AGAAGA	5.61687	0.770722	4.84615	1	1.15402	4.46285	1
GAAGAA	4.86716	0.671663	4.1955	2	0.854474	4.01269	2
AAGAAG	4.29174	0.660342	3.6314	3	0.878069	3.41367	3
TGAAGA	3.95867	0.516819	3.44185	4	0.991471	2.9672	4
GAAGAT	3.42623	0.371741	3.05449	5	0.60352	2.82271	5
AAGAAA	3.55635	1.16015	2.3962	6	1.05244	2.50391	6
GAAGAG	2.64727	0.491786	2.15548	7	0.719154	1.92812	8
GGAAGA	2.62936	0.50618	2.12318	8	0.721032	1.90833	9
CTGAAG	2.24496	0.439994	1.80497	9	0.732098	1.51286	11
GAGAAG	2.24854	0.480449	1.76809	10	0.732883	1.51566	10
AAAGAA	2.8729	1.15234	1.72056	11	0.888027	1.98487	7
AAGATG	2.05633	0.4243	1.63203	12	0.693867	1.36246	14
CAGAAG	2.18049	0.562355	1.61813	13	0.954301	1.22619	16
GAAGAC	1.89636	0.307239	1.58912	14	0.77801	1.11835	20
ATGAAG	1.92024	0.4163	1.50394	15	0.556977	1.36326	13
AGAAAA	2.43835	1.04237	1.39598	16	0.943035	1.49531	12
AGGAAG	2.01515	0.686074	1.32908	17	0.745566	1.26958	15
AAGAGA	1.95844	0.634115	1.32432	18	0.771916	1.18652	18
AGAGAA	2.07364	0.77395	1.29969	19	0.858814	1.21483	17
CAAGAA	1.7167	0.471291	1.24541	20	0.567135	1.14956	19

B

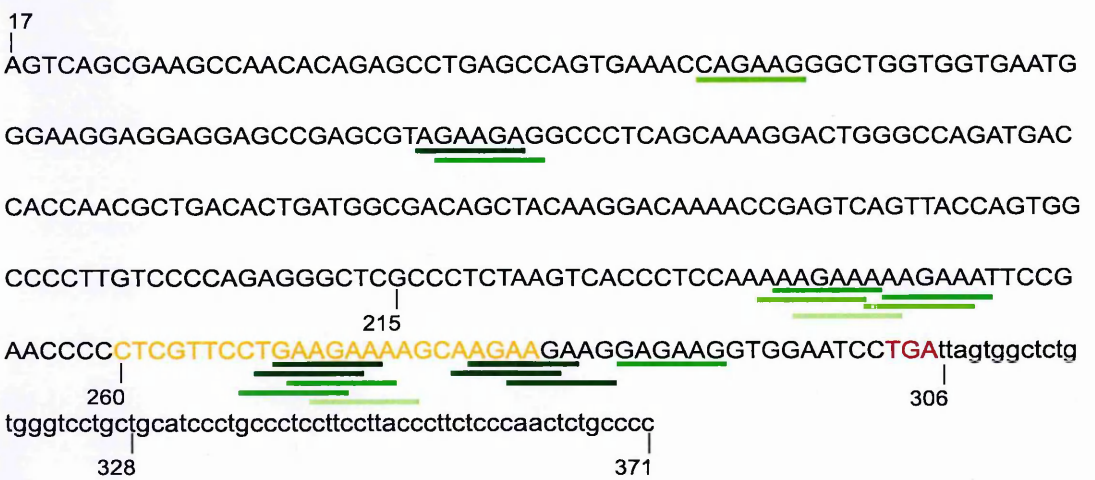


Figure 39
SRSF10 (Tra2beta) putative target sites in the 355nt long enhancer element.
A: List of the 20 most frequent hexamers found in CLIP-tags of SRSF10, as described in Grellscheid et al., 2011; **B**: Sequence composition of the 355 nt long enhancer element; 13 out of the 20 hexamers shown in panel A aligned to the 355 nt long enhancer element; green lines indicate hexamer positions, the color code shown in panel A was respected; numbers indicate the nucleotide positions in respect to the first nucleotide of the mouse beta adducin last exon; stop codon is indicated in red; 24 nt long upstream enhancer element is indicated in orange.

3'splice site and its replacement could lead to the inactivation of splicing which causes a decrease in cleavage and polyadenylation. Similarly, the beta adducin non-canonical enhancer element is positioned in the vicinity of the 3'splice site of the terminal exon. It is well known that terminal-exon definition relies on the crosstalk between splicing and 3'end processing events (Martinson, 2011) (see the Introduction Section). Thus, the non-canonical enhancer element could function in the activation of PAS4 by direct stimulation of terminal exon definition and splicing of the last intron. Preliminary data showed that removal of all introns from the **A1-A23-A4** construct abolished its expression, as no signal was detected by Northern blot analysis (A. Iaconcig and L. Costessi, unpublised data). In the present study we showed the absence of expression of the **A4-A1-A23-A4** construct in HeLa cells, which remained peculiar (Figure 18). As mentioned above, the "proximal" PAS4 is probably inactive due to the absence of the upstream enhancer, which is present downstream to the first PAS4. However, the reason why the distal PAS4 is not active in this context remains unclear. One possible explanation could be that the non-canonical enhancer element is not functional when located far away from the 3'splice site of the terminal exon. Insertion of the non-canonical enhancer element in the proximity of the 3'splice site, upstream of the first PAS4 of the **A4-A1-A23-A4** construct resulted in the activation of the "proximal" PAS4 (data not shown). The possibility that transcription was blocked downstream of the first PAS4 was ruled out since a construct lacking the region downstream to the cleavage site also failed to generate detectable transcripts (Figure 20, construct **A4[**5405-6015**]-A1-A23-A4**).Therefore, it is plausible that the mechanism of the non-canonical enhancer element to activate polyadenylation of the PAS4 is connected to splicing of the terminal exon. This would support the hypothesis that the key factor could be a splicing factor such as SRSF10 (Tra2beta).

Blocking of the translation by cycloheximide showed that transcripts derived from the constructs lacking the non-canonical enhancer element were not subjected to NMD or other cytoplasmic surveillance pathways (Figure 22, **A4** and **INA1-A4** constructs). Instead, these transcripts were inefficiently processed which probably led to their degradation

by the nuclear mRNA quality control pathways (Figure 23). Recently, it has been proposed that the longer is the distance between the natural stop codon and the 3'end of an mRNA, there is increased probability that the mRNA is prone to NMD (Brogna and Wen, 2009). As the enhancer element comprises stop codon, it would be interesting to check for its possible role in mRNA stability regulated by NMD.

In addition to the non-canonical upstream polyadenylation enhancer element, the present study also showed the presence of a non-canonical upstream polyadenylation silencer element which inhibited processing at the mouse beta adducin PAS4. The non-canonical silencer element was 200 nt long and spanned from base 1581 of the beta adducin 3'UTR, respective to the first base of the terminal exon. Removal of this 200 nt silencer element from the **A1-A23-A4** construct led to the three-fold increase in the mRNA levels (Figure 25). The possibility that this element is affecting the half life of the mRNA was ruled out by the mRNA stability assays. In fact, the mRNA stability assays demonstrated that the stability of the mRNA derived from the **A1-A23-A4** construct, thus containing the silencer element, was similar to that of the **A1-A4** construct, which lacks this element (Figure 31).

As observed for the enhancer element, the non-canonical silencer element also had some characteristics that distinguished this USE from other reported ones. First, the non-canonical silencer element represents a long-distance polyadenylation regulatory element. In the minigene context, the distance from the silencer element to the PAS4 is about 800 bases, and in the natural context of the mouse beta adducin gene it is about 4 Kb. Another interesting observation is that USEs are usually enhancers of polyadenylation (Schek et al., 1992; Moreira et al., 1995; Brackenridge and Proudfoot, 2000) and here I described the presence of a silencer of polyadenylation. The presence of an upstream polyadenylation silencer element has also been described for the U1A pre-mRNA and is involved in U1A protein autoregulation (Boelens et al., 1993). This silencer element is ~50 nt long and is located 19 nt upstream of the Hm of the U1A pre-mRNA. The U1A protein binds to the silencer element and directly interacts with PAP inhibiting the polyadenylation reaction, which leads to the degradation of the U1A pre-mRNA and

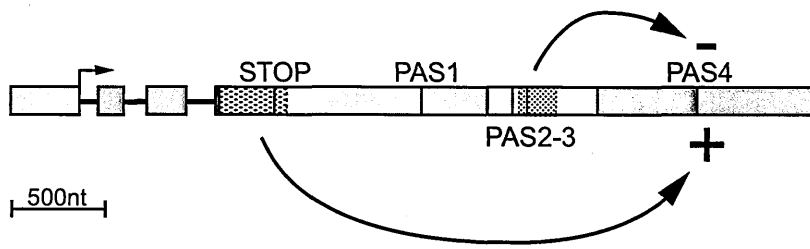


Figure 40
Model of the beta adducin PAS4 activation.
In a tandem array of all three mouse beta adducin PASs, the PAS4 is predominantly used in HeLa cells. The core cis-acting elements are necessary for the correct definition and processing at the PAS4. However, those elements are not enough for the activation of the PAS4. Additional, long-distance non-canonical polyadenylation elements are regulating the usage of the PAS4. The non-canonical long-distance enhancer element is indispensable for the activation of the PAS4 and is located in a 355 nt long region comprising mouse beta adducin stop codon (highlighted with dots). The non-canonical long-distance silencer polyadenylation element is inhibiting the usage of the PAS4 and is located in the 200 nt long region downstream of the PAS2-3 (highlighted with dots). These findings underscore the complex mechanisms regulating the processing at the mouse beta adducin PAS4.

decrease in the levels of the U1A protein (Gunderson et al., 1994; Gunderson et al., 1997). A very recent study revealed a widespread adenosine methylation in mRNAs of mouse brain, which was increased near the stop codons and in 3'UTRs (Meyer et al., 2012). Interestingly, the non-canonical enhancer and silencer elements described here are located around the stop codon and in the 3'UTR of the mouse beta adducin mRNA, respectively. By examining the gene regions containing adenosine methylation in the above mentioned study, I found that the brain-specific beta adducin transcripts contain adenosine methylation and the regions containg this modification overlap with the non-canonical enhancer and silencer elements. It was shown that mRNA adenosine methylation is important for embryonic development of Arabidopsis, oogenesis in Drosophila, nutrient signaling in mammals and some studies connected it with RNA processing. However, the functions of mRNA adenosine methylation are still largely uncharacterized and it would be of interest to check for functional connections between the beta adducin mRNA adenosine methylation and the novel non-canonical polyadenylation regulatory elements.

Taken together, in the presented study I identified novel non-canonical polyadenylation-regulatory elements that enhance or silence 3'end processing of the mouse beta adducin PAS4 (Figure 40). The non-canonical upstream enhancer element is indispensable for the efficient use of the PAS4. On the contrary, the non-canonical upstream silencer element inhibits the use of the PAS4. These results underscore the complex mechanisms regulating polyadenylation of the beta adducin gene.

3.2 Function of the long beta adducin 3'UTR

Usage of the PAS4 leads to the formation of the beta adducin mRNA containing an unusually long 3'UTR (Costessi et al., 2006). The preliminary analyses presented in the last section of the chapter Results were directed to study the role of the long 3'UTR in the mRNA stability and translation efficiency.

Using firefly-luciferease based reporter constructs containg the different 3'UTRs of beta adducin, I showed that the presence of the long mouse beta adducin 3'UTR variant led to a decrease in luciferase activity in all the three cell lines tested, HeLa, MEL and SK-N-BE (Figure 36B). A similar trend was observed in previous studies (Mayr and Bartel, 2009; Sandberg et al., 2008). Performing luciferase assays, they showed that the shorter 3'UTRs led to the production of the higher levels of protein than the longer ones, which is explained by the presence of a higher number of regulatory elements in the long 3'UTRs that could provoke destabilization of the mRNA and translational repression.

Quantification of the mRNA levels from HeLa cells transfected with the luciferase –reporter constructs led to the conclusion that the detected decrease in luciferase activity of the construct bearing the beta adducin long 3'UTR was mostly due to mRNA destabilization and, to a minor extent, translational repression (Figure 36C). However, it should be taken into account that the length of the 3'UTR itself could have influence in the final protein production in HeLa cells, as was noticed by Mayr et al. (Mayr and Bartel, 2009). They analyzed luciferase activity in 15 cell lines using reporter constructs containing short, long and antisense oriented long 3'UTRs. In four of the 15 tested cell lines, the antisense orientation of the long 3'UTRs lost their repressional capacity, influencing the expression of the reporter constructs at the same level as the short ones. Whereas, in the other tested cell lines, including HeLa, intermediate levels of luciferase activity were observed in the presence of the antisense oriented long 3'UTRs.

Preliminary results suggested that the decrease in the luciferase mRNA levels, observed in the presence of the mouse beta adducin long 3'UTR, was probably linked to the presence of the ARE and CPE destabilizing elements. ARE elements are well known in influencing the mRNA stability (von Roretz et al., 2011). Here, it was observed that the deletion of only one ARE from the beta adducin 3'UTR led to an increase in mRNA levels, suggesting that this element has an important role in the regulation of beta adducin mRNA stability. However, further studies are necessary to prove this observation and to identify the potential ARE binding proteins involved in the destabilization of the beta adducin mRNA.

Mutation of the two highly-conserved CPEs located upstream of the beta adducin PAS4 led to an increase in mRNA levels, suggesting a role of these elements in the stability control of the beta adducin mRNA. This observation is peculiar, as CPEs are known to influence translational repression when bound by CPEB in the cytoplasm. A possible explanation for this observation could be that the CPE consensus sequence, **UUUUAU** hexamer, being an **AU**-rich sequence, could also be recognized by members of the ARE-binding protein family. Influence of CPEs in mRNA stability has also been found in p53 3'UTR (Rosenstierne et al., 2008). Mutation of two CPEs present in the p53 3'UTR

led to an increase in mRNA stability and twofold increase in mRNA levels of the reporter gene (Rosenstierne et al., 2008). The role of the CPEs in the stability of the beta adducin mRNA will be addressed in future studies. Additionally, it would be interesting to study the function of CPEs and the CPEB protein in the translational repression of beta adducin mRNA. Upstream of the two CPEs present in the long 3'UTR of the beta adducin mRNA, I also identified one conserved Pumilio-binding element (PBE). The PBE is also involved in translational repression and is usually found close to CPEs. Actually, the position and number of the CPEs in respect to the PBE and Hm influences the efficiency of the translational repression (Pique et al., 2008). Interestingly, CPEB shows dendritic localization in hippocampal neurons, where it has a role in the regulation of local translation of mRNAs containing CPEs (Richter, 2001). Moreover, CPEB has a role in dendritic targeting of several mRNAs, one of which is CAMKIIα mRNA. The major mRNA isoform of beta adducin in brain contains the long 3'UTRs variant and it is expressed at high levels in the hippocampus, cerebellum and neocortex showing dendritic localization (Porro et al., 2010). Therefore, CPEB represents a good candidate for the regulation of the long beta adducin mRNA isoform targeting to dendrites and its local translation.

In addition, translational repression of the long beta adducin mRNA can be caused by miRNA targeting to its 3'UTR. I screened the mouse beta adducin long 3'UTR for the presence of potential miRNA binding sites and observed two interesting candidates, mir-125 and mir-219, which were found to be part of the top 50 most conserved octamer motifs present in the 3'UTR (Xie et al., 2005). The putative target site for mir-125 was located immediately after the mouse beta adducin PAS1. The putative target site for mir-219 represented the longest stretch of conservation found in proximity to the mouse beta adducin PAS23, and overlaped with the potential Hm region. The position of both miRNA binding sites raises the possibility that they could interfere in a tissue-specific manner with polyadenylation at the proximal sites. The presence of the mir-125 and mir-219 putative target sites in the long beta adducin 3'UTR led to the possibility that mir-125 and mir-219 are regulators of the beta adducin mRNA stability in brain. In

addition, they could account for the observed decrese in the luciferase activity of the reporter gene bearing the long beta adducin 3'UTR (Figure 36B). Interestingly, both of these miRNAs are expressed at high levels in mouse brain (Bak et al., 2008). Mir-125 was shown to be involved in the regulation of synaptic function and structure in mouse brain (Edbauer et al., 2010), whereas mir-219, in the regulation of oligodendrocyte differentiation and myelination (Dugas et al., 2010). I checked for the expression of these miRNAs in several mouse tissues and cell lines, performing Northern blot analysis (data not shown). As detected previously (Bak et al., 2008), high levels of mir-219 and mir-125 were observed in mouse brain. The expression of the mir-219 was not observed in any of the other tested tissues (spleen, heart, kidney and lung) and cell lines (HeLa, K562, SK-N-BE and HEK293), whereas mir-125 was expressed at various levels in these tissues and cells but to a lesser extent then in the brain. Therefore, it would be of interest to study the role of these miRNAs in the regulation of beta adducin expression and alternative-polyadenylation site selection in brain.

To summarize, the long beta adducin 3'UTR contains multiple potentially important regulatory elements. Preliminary results suggested the involvement of ARE and CPE elements in the regulation of beta adducin mRNA stability. Further studies will be conducted in order to get deeper insight into the functions of the long beta adducin 3'UTR and the mechanisms regulating its fate.

Chapter 4

Materials and Methods

4.1 Reagents and solutions

Commonly used chemicals were purchased from Sigma-Aldrich, Boehringer-Mannheim, Amersham Biosciences (now GE Healthcare), Merck, Gibco and Invitrogen (now Life technologies).

4.1.1 Standard solutions

Commonly used solutions are listed below:

5X TBE:
450 mM Tris-borate, 10 mM EDTA, stored at RT

TE:
10mM Tris-HCl (pH 7.4), 1mM EDTA (pH 7.4), stored at RT

PBS:
137 mM NaCl, 10 mM Na_2HPO_4, 2.7 mM KCl, 2 mM KH_2PO_4, pH 7.4, stored at 4°C

10X MOPS:
200 mM MOPS (pH 7.0), 50 mM CH_3COONa (pH 8.0), 10 mM EDTA (pH 8.0), stored at 4°C

20X SSC:
3 M NaCl, 0.3 M sodium citrate (dihydrate), pH 7.0, stored at RT

5X loading dye for agarose gels:
40% (w/v) sucrose, 0.25% (w/v) bromophenol blue, stored at RT

5X formamide-loading buffer:
1 mg/ml xylene cyanol, 1 mg/ml bromophenol blue, 80% (w/v) deionized formamide, 10 mM EDTA (pH 8.0), stored at -20°C

1X RNA sample buffer for denaturing agarose gels:
50% (v/v) deionized formamide, 1X MOPS, 6% (v/v) formaldehyde, 5 ml ethidium bromide (1 mg/ml) per 1 ml sample buffer, stored at -20°C

R-loop buffer:
80% (v/v) deionized formamide, 40 mM PIPES (pH 6.4), 1 mM EDTA (pH 8.0), 400 mM NaCl, prepared prior to use or stored at -80°C

10X OPA buffer:
100 mM Tris-HCl (pH 7.5), 100 mM Mg acetate, 500 mM K acetate, stored at -20°C

TSS:
10% (w/v) PEG (MW=4000), 5% (v/v) DMSO, 35 mM $MgCl_2$, LB broth (pH 6.5), prepared prior to use or stored at 4°C for a short period

Lysis buffer:
50 mM glucose, 20 mM Tris (pH 8.0), 10 mM EDTA, stored at RT

SDS buffer:
1% SDS, 0.2 M NaOH, stored at RT

4.1.2 Enzymes

The following enzymes were used: restriction endonucleases (New England Biolabs), Taq DNA polymerase (New England Biolabs, Roche), T4 DNA ligase (Roche), T4 DNA polymerase (Roche), Klenow fragment of DNA polymerase (New England Biolabs), T4 polynucleotide kinase (New England Biolabs), calf intestinal alkaline phosphatase (CIP) (Roche), Pfu DNA polymerase (Stratagene), M-MLV reverse transcriptase (Invitrogen), RNase free-DNase (Promega), T7 RNA polymerase (Stratagene), RNase A (Sigma Chemicals), RNase T1 (Sigma-Aldrich), proteinase K (Sigma-Aldrich).

All enzymes were used following the manufacturer's instructions, except for the restriction endonucleases and CIP, which were all used with OPA buffer, instead of the manufacturer's recommended buffer.

4.1.3 Radioactive isotopes

Radioactive nucleotides [α-^{32}P]dCTP and [α-^{32}P]UTP were purchased from Amersham Bioscience.

4.1.4 Synthetic oligonucleotides

Synthetic oligonucleotides were purchased from Sigma-Aldrich or MWG biotech.

4.1.5 Biological reagents

The E.coli DH5α and Dam- strains were used for the amplification of the plasmids. For the transfection experiments HeLa (human cervical cancer cells), SK-N-BE (human neuroblastoma cells) and MEL (murine erythroleukemia cells) cell lines were used.

4.2 PCR

For preparative PCRs Taq Polymerase supplied by Roche was used, whereas New England Biolabs Taq Polymerase was used for analytical PCRs. The final concentration of the components in the PCR reaction mix were: 1X corresponding PCR buffer (10X stock), 100 nM dNTPs (stock 5 mM), 3.3 ng/µl forward and reverse primers (100 ng/µl stock) and 0.025 U/µl of Taq polymerase (5 U/µl stock). The following termocycler steps were used for the amplification of a ~0.5-1 Kb long product: first step was denaturation at 94°C for 3 min, then 30 cycles of denaturation at 94°C for 30 sec, annealing at 56°C for 30 sec and elongation at 72°C or 68°C (Taq Roche and Biolabs, respectively) for 30

sec, and the last step was elongation at 72°C or 68°C (Taq Roche and Biolabs, respectively) for 5 min. In the case the fragments were longer than 1 Kb, for each additional 500 bp the elongation step was prolonged for 30 sec.

4.3 Enzymatic modifications of DNA

4.3.1 Blunt end formation

In order to form blunt ends, PCR products were treated with Klenow or T4 DNA polymerase. Briefly, after the PCR reaction, the Taq polymerase was heat inactivated at 98°C for 10 min. Subsequently, 2.5 µl of $MgCl_2$ and 1 µl of Klenow were added in a 46 µl of the PCR and the reaction was incubated at room temperature (RT) for 15 min. The Klenow enzyme was then inactivated at 80°C for 10 min, followed by phosphorylation of the newly formed blunt ends by T4 polynucleotide kinase: 14 µl of ATP [10mM], 1.4 µl of T4 kinase buffer, 2.5 µl of EDTA [3mM] and 1 µl of T4 polynucleotide kinase were added to the reaction, following incubation at 37°C for 1 hour. Blunt ended and phosphorylated PCR fragment were then gel purified and used for the ligation.

4.3.2 Restriction

For analytical purpose, usually 100-300 ng of plasmid DNA were digested with restriction endonucleases, in a 15 µl of final reaction volume. For preparative samples, 2 µg of plasmid DNA or 10-20 µl of PCR product were digested in 50 µl of final reaction volume. 10X OPA buffer was used for all the reactions, in final concentrations optimal for the used enzymes. Reactions were carried for at least one hour at 37°C for most of the enzymes. Depending on the purpose, the enzymes were heat inactivated, removed by gel purification or by phenol-chloroform purification.

4.3.3 Phosphatase

After digestion and heat inactivation of restriction enzymes, linearized plasmid DNA was treated with CIP in order to remove phosphates and to abolish self-ligation. 0.5 µl of CIP (1U/µl stock) was added for 1 µg of plasmid. Incubation was carried at 37°C for 1 hour. After the reaction, CIP was inactivated at 65°C for 20 min or removed through gel purification of the plasmid.

4.3.4 Ligation

Ligation was carried out with 20-30 ng of digested plasmid and 3-6 folds excess of insert, in a 15-20 µl of final volume containing T4 ligase buffer (10X stock) and 1U of T4 ligase. Two controls were always performed to determine the quality of the vector and ligation reaction: CUT (without ligase and insert, to check for the efficiency of digestion) and SELF (without insert, to check for the plasmid self-ligation). Ligation reactions were incubated at RT for 2-5 hours. Half of the reaction was used for transformation and the other half was incubated ON at RT and used for transformation the day after.

4.4 Transformation

4.4.1 Preparation of competent bacteria (E.coli DH5α)

A single bacterial colony was grown ON in 5 ml of LB at 37°C with agitation. 5 ml of ON inoculum was transferred into a 200 ml LB and incubated at 37°C with agitation until the absorbance at 600nm reached 0.3-0.4 units (optimal 0.35). Next, bacteria were centrifuged at 3000 rpm for 5 min at 4°C. The supernatant was discarded and the pellet was gently resuspended in TSS solution. Aliquots of 400 µl were transferred into 1.5 ml eppendorf tubes, immediately frozen in liquid nitrogen, and stored at -80°C. Competence

was tested transforming bacteria with 0.1 ng of the pUC19 vector. Generally, competence was ~10^7 colonies per µg of plasmid DNA.

4.4.2 Transformation of competent bacteria (E.coli DH5α)

Half of the ligation reaction or 30 ng of plasmid (in the case of retransformation) was incubated for 30 min in ice with 70 µl of competent bacteria. Next, bacteria were heat-shocked at 42°C for 2 min and placed on ice for 1 min, 70 µl of pre-warmed LB was added to the bacteria and incubated for 30 min at 37°C. Afterwards, bacteria were plated in pre-warmed LB Petri plates containing the corresponding antibiotics. When necessary, 70 µl of IPTG (100 mM) and 35 µl of X-gal (4%) were added to the plate. Plates were incubated ON at 37°C to allow for bacteria growing and subsequently removed and stored at 4°C.

4.5 DNA purification

4.5.1 Band purification

For cloning, linearized vectors and inserts or blunt-ended PCR products were separated on agarose gel of appropriate concentration. Bands corresponding to the desired DNA size were cut and DNA extracted from the gel slices using the Qiaquick gel extraction kit (cat. No. 28706) from Qiagen, according to the manufacturer's instructions.

4.5.2 PCR clean-up

PCR products used as templates for the *in vitro* transcription reactions were purified using NucleoSpin Gel and PCR clean-up kit from Macherey-Nagel, according to the manufacturer's instructions for the PCR clean-up protocol.

4.5.3 Phenol-chloroform purification

After treatments with different DNA modification enzymes, DNA was purified with phenol-chloroform. One volume of phenol-chloroform was added to the DNA sample, which was vortexed and centrifuged for 2 min at 13200 rpm. The supernatant was transferred into a clean 1.5 tube containing 2 volumes of 100% ethanol with 1/10 (v/v) of sodium acetate (3 M, pH 5.0). The sample was vortexed and left in ice for 10 min. Subsequently, the sample was centrifuged at 13200 rpm for 15 min. The supernatant was removed, the pellet was washed with 500 µl of 70% ethanol and resuspended in dH_20.

4.6 Agarose gel electrophoreses of DNA

Different concentrations of agarose, from 0.6% to 2% (w/v) were used for the preparation of gels, according to the expected size of DNA fragments. Agarose was melted in 1X TBE, which was also used as a running buffer. Gels contained ethidium bromide in order to visualize DNA with UV transiluminator. To determine the size of the bands, the 1 Kb molecular weight marker was used, purchased from Invitrogen (cat.No. 15615-024).

4.7 DNA preparation

4.7.1 Mini-prep

Small-scale preparations of plasmid DNA from bacterial cultures (mini-prep) were performed in order to screen for the positive clones. Each of the selected colonies was picked with a toothpick and introducesd into a tube containing 2.5 ml LB and the corresponding antibiotic (1/1000 v/v, stock 10 mg/ml). The inoculum was incubated at 37°C ON with agitation. Afterwards, the inoculum was transferred into a 1.5 ml-tube and centrifuged for 5 min at 13200 rpm. The supernatant was removed and the pellet was resuspended in 90 µl lysis buffer. The cells were lysed by adding 200 µl of SDS-buffer and mixed by inversion 4-5 times. Then, 150 µl of NaAc (3 M, pH 3.2) were added, mixed 4-5 times by inversion and left for 2 min on ice. Subsequently, the sample was centrifuged for 15 min at 13200 rpm. The supernatant was purified with one volume of phenol-chloroform, DNA was precipitated in two volumes of 100% ethanol for 5 min on ice and washed with 70% ethanol. The DNA was resuspended with 30 µl dH_20 and 2 µl RNase A (10 mg/ml stock) and incubated for about 10 min at 37°C. An aliquot of the DNA samples was then subjected to restriction enzyme digestions. Digested DNAs were separated on agarose gel and the size of the bands was visualized on the UV transiluminator.

4.7.2 Commercial mini-prep

For the purposes of sequencing, cloning and transfection, plasmid DNA was prepared using "Wizard plus SV minipreps DNA purification system" from Promega (cat. No. A1460). DNA preparation was performed according to the manufacturer's instructions.

4.7.3 Midi-prep

Large scale preparation of plasmid DNA from bacterial cultures (midi-prep) was per-

formed using "Plasmid purification midi kit" from JetStart (cat. No. 210025), according to the manufacturer's instructions.

4.8 Measurment of nucleic acids concentration and purity

DNA and RNA concentration were measured using a nano-drop spectrophotometer. As an indication of sample purity, the ratio of the absorbencies at 260 nm and 230 nm was used, considering sufficiently pure samples with the ratio 1.7 and above. The concentrations were estimated measuring the absorbance at 260 nm.

4.9 Sequencing

PCR amplified products were subcloned into pUC19 and sequenced prior to their transfer into the final vector. Sequencing reactions were performed by "BMR Genomics" based in Padova, Italy.

4.10 Mutagenesis

4.10.1 Quick-change PCR mutagenesis protocol

The quick-change PCR mutagenesis protocol was used to introduce point mutations into the desired fragments (QuickChange Site-Directed Mutagenesis Kit protocol, Stratagene). The amplification was performed with a pair of complementary primers

containing the mutation flanked by about 15 unmodified nucleotides at each side. 150 ng of the plasmid to be mutated was used as template in a 30 µl PCR reaction. The high fidelity Pfu DNA polymerase (Stratagene) was used. The PCR steps were as follows: denaturation at 95°C for 30 sec, 18 cycles of denaturation at 95°C for 30 sec, annealing at 55°C for 1 min and elongation at 68°C for 4 min, and one last step of elongation at 68°C for 7 min. The original plasmid used as a template was digested with Dpn I (it does not recognizes the DpnI sites of Taq-polymerase derived DNA), leaving only the mutated plasmid in the PCR reaction. 25 µl of the PCR reaction were digested at 37°C for one to 4 hours with 20 U of the Dpn I enzyme. 14 µl of the reaction were used to transform bacteria as described above.

Generally, the template DNA was pUC19 plasmid containing the region with the desired sequence to mutate. The mutated pUC19 plasmid was then sequenced to check for the presence of the desired mutation and the region of interest was cloned in a final vector.

4.10.2 Two-steps PCR technique

In order to create deletions, a two-steps PCR technique was used. In the first step two PCR reactions were performed: the first one was performed with a forward (FWD) external and a reverse (REV) internal primer and the second PCR with a FWD internal and a REV external primer. The PCR fragments from these two reactions were gel purified and used as templates for the second step of the PCR. The external FWD and REV primers were used for this PCR. The internal primers were designed to overlap in the ~15 nt long region upstream and downstream of the desired deletion, thus allowing the annealing and joining of the two fragments in the second step of the PCR reaction. The external primers contained two pairs of restriction enzyme sites at their ends (external and internal). The PCR fragment obtained in the second step was digested in the positions of the external restriction enzyme sites and subcloned into pUC19. Sequencing of the insert inside the PUC19 was performed in order to check for the presence of the desired deletion. Next, the mutated fragment was cut and cloned in the final vector.

4.11 Generation of minigene constructs

4.11.1 Generation of chimeric mouse beta adducin minigene constructs

The **A1-A23-A4** and **A4-A1-A23-A4** constructs were previously generated in the laboratory using the pBS SV40 alpha globin minigene. The pBS SV40 alpha globin minigene construct derived from pBS II KS plasmid (Stratagene) which contains 352 bp of the SV40 promoter and the genomic sequence human alpha globin gene including all three exons with their corresponding introns (1071 bp). The last 344 bp of the alpha globin third exon containing the PAS were removed by digesting the construct with BstE II (located in the coding part of the third exon) and Sal I (located in the pBS II KS polylinker) and replaced with a linker (linker2) containing several unique restriction sites. This construct was called pBS SV40 alpha globin linker2. The mouse beta adducin genomic regions containing PASs were called **A1**, **A23** and **A4**, and are described in Figure 16A. These regions were inserted into the correct sites of the linker2 to create the **A1-A23-A4** construct (Figure 16B). The additional **A4** region was cloned in the sites present upstream of the **A1** region in the **A1-A23-A4** construct to create **A4-A1-A23-A4** construct (Figure 18A). The **A1-A23-A4** constructs containing deletions of the polyadenylation signal elements (Figure 17) were created using Quick-change PCR mutagenesis method. All of these constructs were created prior to my arrival at the lab.

Two additional mouse beta adducin exons and introns were inserted into the BstE II position of the **A1-A23-A4** and **A4-A1-A23-A4** constructs to generate **EX-A1-A23-A4** and **EX-A4-A1-A23-A4** constructs (Figure 19A). The region starting at base 68 of exon 14 to base 1910 (in respect to the first base of the exon 14) in the intron 16 was amplified from mouse genomic DNA using the two-steps PCR technique in order to remove a BstE II site present in intron 14.

The following primers were used:

First PCR reaction

FWD external primer:
mB-Add Ex14 68 dir BstEII - 5' CACGGTGACCAACCCCTTCAGCCAACTCAC 3'
REV internal primer:
mB-Add 803 rev BstEII mut - 5' TCCAGAGGGTCACGTGGGGCTTGTTCT 3'

Second PCR reaction

FWD internal primer:
mB-Add 778 dir BstEII mut - 5' AGAACAAGCCCCACGTGACCCTCTGGA 3'
REV external primer:
mB-Add 1910 rev BstBI - 5' CAGTTCGAAGAAGGACTGTGAGAAGGAAA 3'

Both PCR reactions were gel purified and used as templates for the joint PCR reaction with both external primers. The amplified PCR product, containing a BstE II site on its 5'end and a BstB I site on its 3'end, was blunt-ended and subcloned into a pUC19 vector for sequencing.

The second region contained the last 846 bp of intron 15 and the first 16 bp of exon 16 and was amplified by PCR from mouse genomic DNA using the following primers:

FWD:
mB-Add 8050 dir BstBI - 5' CAGTTCGAAACTCAAATCGTCACCTCTCACAC 3'
REV:
mB-Add Ex14 8888 rev BstEII-5'CACGGTCACCTTCTGTCTTCTTGGTATCTAGGA3'

The BstE II site was at the 5'end of the amplified PCR product, whereas the BstB I site was present on its 3'end. This PCR product was blunt-ended and subcloned into pUC19 for sequencing.

Finally, the first region was cut from the pUC19 vector with BstB I and Hind III, gel purified and joined to the second region through the internal BstB I site, leaving the BstE II sites from both ends of the newly formed larger region. Thus, the formed new region contained the last 87 bp of exon 14, all intron 14 (957 bp) with the mutated BstE II site, all exon 15 (126 bp), the first 674 bp of intron 16 joined with BstB I with the last 864 bp

of intron 16 and the first 16 bp of exon 16. This region was cut from the pUC19 plasmid with BstE II, gel purified, CIP treated and ligated into the BstE II site of the **A1-A23-A4** and **A4-A1-A23-A4** constructs to generate the **EX-A1-A23-A4** and **EX-A4-A1-A23-A4** constructs (Figure 19A).

For the generation of **A4[**5405-6015**]-A1-A23-A4** and **EX-A4[**5405-6015**]-A1-A23-A4** constructs (Figure 20A) , the short **A4** region, spanning from the base 5405 to the base 6015, in respect to the first one of the mouse beta adducin last exon, was cloned inside the **A1-A23-A4** and **EX-A1-A23-A4** constructs, upstream of the **A1** region. This short **A4** region was amplified using the following primer pair:

FWD:
BR A4 dir 1 - 5' GGGGTACCGCGGCCGCCATATGAGTGTCTCGTAAGTAATGCT 3'
(Kpn I – Not I – Nde I tail)

REV:
mAdd2BRshort rev -5'CGGGATCCCCGCGGGAATTCCCTACATTTCCATACATGA3'
(BamH I – Sac II – EcoR I tail)

The PCR fragment was digested at the external restriction enzyme sites (Kpn I and BamH I), subcloned into pUC19 which contained the mutated EcoR I and Nde I sites, and sequenced. The short **A4** region was cut from the pUC19 plasmid with Not I and Sac II, gel purified and cloned into the Not I and Sac II positions upstream of the A1 region in the **A1-A23-A4** and **EX-A1-A23-A4** constructs.

The **A4** construct (Figure 21A) was generated by placing the **A4** region into the Nde I and EcoR I sites of the linker2 in the construct pBS SV40 alpha globin linker2.

The **A1** region was removed with Spe I and Xba I digestion from the **A1-A23-A4** construct. The digested construct was then treated with Klenow and religated without purification in order to obtain the **A23-A4** construct (Figure 21A).

The **A1** region was added to the **A4** construct into the Spe I and Xba I linker positions to obtain the **A1-A4** construct (Figure 21A).

To generate **INA1-A4** construct (Figure 21A), the **A1** region was cut from the **A1-A23-A4**

construct with Spe I and Xba I, treated with Klenow, gel purified and inserted into the **A4** construct which was cut with Spe I and EcoR V, treated with Klenow and gel purified. The presence of the **A1** region in the antisense orientation was validated by restriction enzyme digestion which cuts the plasmid once only in the antisense **A1** region (Afl II). The different subregions of the **A1** region were PCR amplified. The primer pairs contained restriction sites necessary for the cloning strategy at their 5'end. The external sites were BamH I and Pst I and the internal ones Spe I and Xba I. The PCR products were digested with BamH I and Pst I. After heat inactivation of the enzymes, the PCR fragments were subcloned into pUC19 in the BamH I and Pst I position. The pUC19 plasmids containing the different subregions of the **A1** region were sequenced. Then, the subregions of the **A1** region were cut out with Spe I and Xba I, gel purified and inserted into the final vector. The final vector was prepared by digestion of the **A1-A4** construct to remove the whole **A1** region, with Spe I and XbaI. The enzymes were heat inactivated and the cut plasmid was CIP treated. The band corresponding to the **A1-A4** construct without the **A1** region was then cut from the gel and purified, and used for the ligation reactions with the different subregions of the **A1** region. The primer pairs used for the amplification of each of the **A1** subregions, in the order they appear in the figures, are listed below:

A1[17-305]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
SPA1-STOP_rev – 5' AACTGCAGTCTAGATCAGGATTCCACCTTCTCCTT 3'

A1[284-887]:
SPA1-268UP dir – 5'CGGGATCCACTAGTGAAGGAGAAGGTGGAATCCTGA 3'
SPA1-HALF_rev – 5'AACTGCAGTCTAGACTCAGGGTTAGTGTTTGGAA 3'

A1[868-1416]:
SPA1-HALF_dir – 5' CGGGATCCACTAGTTTCCAAACACTAACCCTGAG 3'
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[17-887]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
SPA1-HALF_rev – 5'AACTGCAGTCTAGACTCAGGGTTAGTGTTTGGAA 3'

A1[284-1416]:
SPA1-268UP dir – 5'CGGGATCCACTAGTGAAGGAGAAGGTGGAATCCTGA 3'
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[215-1416]:
SPA1-199UPdir – 5'CGGGATCCACTAGTGCCCTCTAAGTCACCCTCCA 3'
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[260-1416]:
SPA1-244UPdir – 5' CGGGATCCACTAGTCTCGTTCCTGAAGAAAAGCA 3'
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[17-531]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
SPA1_MscIrev – 5' AACTGCAGTCTAGATGGCCATGGGAATCAGGAAG 3'

A1[17-471]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-60rev – 5' AACTGCAGTCTAGATATGGACAAGACCCCTGGTT 3'

A1[17-420]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-111rev – 5' AACTGCAGTCTAGATGAGTGTGATCCCAGGACCT 3'

A1[17-380]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-151rev – 5' AACTGCAGTCTAGATATGTTCCAGGGGCAGAGTT 3'

A1[17-371]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-160rev – 5' AACTGCAGTCTAGAGGGGCAGAGTTGGGAGAAGG 3'

A1[17-351]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-180rev – 5' AACTGCAGTCTAGAGTAAGGAAGGAGGGCAGGGA 3'

A1[17-328]:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
s1+2SPA1-203rev – 5' AACTGCAGTCTAGAAGCAGGACCCACAAGAGCCAC 3'

A1[260-328]:
SPA1-244UPdir – 5' CGGGATCCACTAGTCTCGTTCCTGAAGAAAAGCA 3'
s1+2SPA1-203rev – 5' AACTGCAGTCTAGAAGCAGGACCCACAAGAGCCAC 3'

A1[215-380]:
SPA1-199UPdir – 5'CGGGATCCACTAGTGCCCTCTAAGTCACCCTCCA 3'
s1+2SPA1-151rev – 5' AACTGCAGTCTAGATATGTTCCAGGGGCAGAGTT 3'

All constructs containing the different deleted versions of the putative upstream and downstream enhancer elements, in the **A1** region (Figure 29A) were performed using the two-steps PCR technique. The **A1** regions containing the deletions were subcloned into pUC19, sequenced and transferred to the final vector as explained above. Primer pairs and templates used for PCR reactions are listed below for each deletion.

A1[Δ260-283] (deletion of the upstream enhancer element):
Template: **A1** region in pUC19
FWD external primer:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
REV internal primer:
dUPrev - 5'TTCTCCTTCGGGGTTCGGAATTTCTTTTTCTTTTTGG 3'
FWD internal primer:
dUPdir - 5' AACCCCGAAGGAGAAGGTGGAATCCTGATTAG 3'
REV external primer:
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[Δ306-328] (deletion of the downstream enhancer element):
Template: **A1** region in pUC19
FWD external primer:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'
REV internal primer:
dDOWNrev - 5' CAGGGATGCTCAGGATTCCACCTTCTCCTTC3'
FWD internal primer:
dDOWNdir - 5'TCCTGAGCATCCCTGCCCTCCTTCCTTACCCT 3'
REV external primer:
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

A1[Δ260-283&Δ306-328]-A4 (deletion of upstream and downstream enhancer elements):
Template: **A1[Δ260-283]** in pUC19
Primer pairs as for the deletion of only downstream enhancer element.

A1[Δ260-328]-A4 (deletion of 69 bases-long putative enhancer element):
Template: **A1[Δ260-283&Δ306-328]-A4** in pUC19
FWD external primer:
SP A1 dir 1 – 5' CGGGATCCACTAGTAAGTCAGCGAAGCCAACACA 3'

REV internal primer:
d69rev - 5' CAGGGATGCGGGGTTCGGAATTTCTTTTTCT 3'
FWD internal primer:
d69dir - 5'AACCCCGCATCCCTGCCCTCCTTCCTTACCCT 3'
REV external primer:
SP A1 rev 1 – 5' AACTGCAGTCTAGAGTGATGGGCTTGGGAATGC 3'

The different subregions of the **A23** region (Figure 25A) were PCR amplified using primer pairs containing at their 5'end the restriction sites necessary for the cloning strategy. The external restriction sites were Kpn I and BamH I, and the internal ones Cla I and EcoR V. The cloning strategy was similar to that one described above for the A1 region. The amplified **A23** subregions were subcloned into pUC19 using the external restriction sites, sequenced and transferred into final vector using internal ones. The final vector was prepared by removing the **A23** region from the **A1-A23-A4** construct, after digestion with Cla I and EcoR V. Next, the vector was gel purified and used for the ligation reactions with the **A23** subregions. The primer pairs used for the amplification of the **A23** subregions are listed below.

A23[1416-1781**]**:
SP A2A3 dir 3 – 5' GGGGTACCATCGATTCTTGCATCCTGAAGCTTG 3'
m-SP A2A3 short rev 2 – 5' CGGGATCCGATATCAGGCAGGACTTGGCTCAGAC 3'

A23[1416-1581**]**:
SP A2A3 dir 3 – 5' GGGGTACCATCGATTCTTGCATCCTGAAGCTTG 3'
SPA23short2rev – 5' CGGGATCCGATATCGAGAAACGCTCTTTGAAGAC 3'

Mutations of the two CPEs present in the **A4** region of the **A1-A23-A4[**mutCPE**]** construct (Figure 37A) were created using Quick-change PCR mutagenesis method. The plasmid used for mutagenesis was a pUC19 vector (with mutated EcoR I and Nde I sites) containing the **A4** region. The mutated **A4** region was sequenced, cut from pUC19 with EcoR I and Nde I (present at the ends of the **A4** region), gel purified and cloned into the EcoR I and Nde I positions of the final vector. The final vector was prepared by

replacing the wt **A4** region from the **A1-A23-A4** construct with the **A4** fragments containing the CPE mutations, by digesting at the EcoR I and Nde I site. The primers used for the Quick-change PCR mutagenesis are listed below (mutations are in lower case).

Mut2xCPEdir – 5' GCTCTTGGATgTcATTTTGCAATcTgATATATATACAT 3'
Mut2xCPErev – 5' GTATATATATcAgATTGCAAAATgAcATCCAAGAGCCT 3'

Deletion of the ARE (TATTTAT sequence) from the **A4** region present in the **A1-A23-A4[**ΔARE**]** construct (Figure 38A) was created using two-steps PCR technique. The PCR A4 fragment containing deletion was digested at the external restriction enzyme sites present at its ends (Kpn I and BamH I), subcloned into a pUC19 vector containing the EcoR I and Nde I sites mutated, and sequenced. The **A4** region containing the ARE deletion was then inserted into the **A1-A23-A4** construct (lacking the **A4** region) as described above for the CPE mutations. The primer pairs used for the two-steps PCR reactions are listed below.

FWD external primer:
BR A4 dir 1 - 5' GGGGTACCGCGGCCGCCATATGAGTGTCTCGTAAGTAATGCT 3'
REV internal primer:
dARErev – 5' CCATATAAAACCAACCCTATATATTATATATGTA 3'
FWD internal primer:
dAREdir - 5' TATATAGGGTTGGTTTTATATGGTTTTCTGTT 3'
REV external primer:
BR A4 rev 1 – 5' CGGGATCCCCGCGGGAATTCACCACTTTTCACCTTATTCT 3'

4.11.2 Generation of reporter constructs

Spleen- and brain-specific mouse beta adducin 3'UTRs were amplified by PCR. The template used for amplifying the spleen-specific 3'UTR was mouse genomic DNA. For the amplification of the long brain-specific 3'UTR, the template was a RIKEN FANTOM™ clone containing the mouse brain full-length cDNA, purchased from K.K. DNAFORM, Japan (clone ID: M5C1070K15). Long range Taq DNA polymerase (Roche) was used for amplification of the brain-specific 3'UTR, whereas for the spleen-specific 3'UTR amplification it was used Taq DNA polymerase (Roche). The primer pairs contained Xba I sites at their 5'ends. PCR amplicons were blunt-ended and subcloned into pUC19 for sequencing. The inserts were then cut, gel purified and cloned into the Xba I-linearized pGL3-control vector (Promega). Thus, each of the 3'UTRs were placed downstream of the stop codon of the reporter firefly luciferase gene. The SV40 late polyadenylation signal of the reporter vector was present downstream of the inserted 3'UTRs, which assured efficient 3'end processing of the reporter gene. The final constructs were named **pGL3-A1**, **pGL3-A1-A23** and **pGL3-A1-A23-A4** and are schematically represented in Figure 36A. The primer pairs used for the amplification of the mouse beta adducin 3'UTRs and the sizes of each 3'UTR are listed below.

Short spleen-specific 3'UTR (present in the construct **pGL3-A1**):
size: 776 nt, spans from the 303 nt to the 1078 nt in respect to the first nt of the mouse beta adducin last exon
mBeta-adducin STOP dir Xba – 5' TGCTCTAGATGATTAGTGGCTCTGTGGGTC 3'
mBeta-adducin A1 rev Xba – 5' TGCTCTAGATTCTTAAGCTACAGCTGGGT 3'

Short spleen-specific 3'UTR (present in the construct **pGL3-A1-A23**):
size: 1332 nt, spans from the 303 nt to the 1634 nt in respect to the first nt of the mouse beta adducin last exon
mBeta-adducin STOP dir Xba – 5' TGCTCTAGATGATTAGTGGCTCTGTGGGTC 3'
mBeta-adducin A2-3 rev Xba – 5' TGCTCTAGACCGATGAGCTAAGAAACTTCCA 3'

Long brain-specific 3'UTR (present in the construct **pGL3-A1-A23-A4**):
size: 5629 nt, spans from the 303 nt to the 5931 nt in respect to the first nt of the mouse beta adducin last exon
mBeta-adducin STOP dir Xba – 5' TGCTCTAGATGATTAGTGGCTCTGTGGGTC 3'
mADD2-A4_rev_Xba – 5' TGCTCTAGAGGGAAGCCAAGCACATCTGG 3'

4.12 Cell culture maintenance

Cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX, D-glucose and Pyruvate, supplemented with 10% of Fetal Calf Serum (FCS) and 100 mg/ml of Normicin (antibiotic/anti-mycotic). Cells were cultured in an incubator at 37°C, 5.5% of CO_2, in a humidified atmosphere. Tests for the presence of mycoplasma in cell cultures were performed frequently. Adherent cells (HeLa, SK-N-BE) were maintained in 100 mm Petri dish, whereas cells grown in suspension (MEL, K562) were maintained in 25cm^2 flasks. To pass the cells grown in flasks, cells were transferred into 15 ml Falcon tubes and centrifuged. The medium was removed, and the pelleted cells were washed twice with PBS, centrifuged and resuspended in 10 ml pre-warmed medium. 2-3 ml of the suspension were placed in a new flask containing 8 ml of medium. To pass the cells grown in Petri dishes, medium was removed, cells were washed twice with PBS and 1 ml of 0.1% (w/v) trypsin was added. After 5 min of incubation at 37°C, 5 ml of pre-warmed medium were added and cells were transferred into a 15 ml Falcon tube and centrifuged for 5 min at 1000 rpm. Pelleted cells were resuspended in 10 ml of medium, of which 2-3 ml were plated on a new 100 mm dish containing 8 ml of medium. Cell stocks (resuspended in FCS with 10% DMSO) were made and kept at -80°C or liquid nitrogen.

4.13 Cell culture transient transfections

Transient transfections of cells were performed using Lipofectamine 2000 reagent (Invitrogen), according to the manufacturer's instructions. For the purpose of transcript analysis, transfections of HeLa cells were carried out in 35mm Petri dish or 6-well plates. 5x10^5 cells in 2 ml of antibiotic free DMEM supplemented with 10% FCS were plated and after 24 hours were transfected (at 90% confluency). 10 µl of Lipofectamine 2000

reagent were diluted in 250 µl of Optimem medium (Gibco) for each DNA sample. After 5 min of incubation at RT°C, these 260 µl of the diluted transfection reagent were added to the DNA sample which was previously diluted in 250 µl Optimem. The DNA mix contained 3 µg of the minigene construct, 500 ng of the plasmid used for checking transfection efficiency (pEGFP-C2 or phRG-TK vector containing renilla luciferase gene with the fibronectin intron inserted, created in the lab) and 500 ng of the plasmid expressing SV40 T-antigen (to stimulate SV40 promoter activity). The DNA samples were incubated with the transfection reagent for 20 min at RT°C to allow the formation of DNA-liposome complexes. Cells were washed twice with PBS and 1.5 ml of Optimem was added. 4-6 hours after the transfection, the cells were washed twice with PBS and antibiotic-free DMEM supplemented with 10% FCS was added. The cells were harvested for the RNA extraction 16-18 hours after the transfection.

For the Luciferase assays, transfections were carried out in 24-well plates. The procedure was the same as described above. $1.5x10^5$ cells in 0.5 ml of antibiotic free DMEM supplemented with 10% FCS were plated. The amount of DNA and Lipofectamine 2000 were adjusted to the 24-well plates, as recommended by the manufacturer's instructions. 2 µl of Lipofectamine 2000 were used for each sample. 300 ng of pGL3 constructs were cotransfected with 150 ng of the phRG-TK renilla plasmid.

4.14 RNA extraction from cell cultures and tissues

Total RNA was extracted from cells in culture and tissues using EuroGold TriFast solution from EuroClone (cat. No. EMR507100), according to the manufacturer's instructions.

4.15 *In vitro* transcription

T7 RNA polymerase was used (Stratagene) for the in vitro synthesis of RNAs. Typical in vitro transcription reaction mix contained: 1X T7 buffer (Stratagene, 5X stock), 1.5 mM NTPs (stock 15 mM), 10 mM DTT (stock 100 mM), 1.25 U/µl T7 RNA polymerase (50U/µl stock) and 0.5 µg of DNA template in 20 µl final volume. The reaction was incubated for 2 hours at 37°C, then 5 U of DNAse was added and incubated for another 30 min. Afterwards the reaction was phenol-chloroform purified (Section 4.17.3) and the synthesized RNA was separated on agarose or denaturing agarose gels (Section 4.16.1 and 4.16.2) to check quality and size.

In order to radioactively label RNAs (hot RNAs), α-^{32}P]UTP was added to the in vitro transcription reaction. The reaction mix contained: 1X T7 buffer (Stratagene, 5X stock), 0.75 mM ATP, CTP and GTP (stock 15 mM), 0.1 mM UTP (stock 1mM), 10 mM DTT (stock 100 mM), 1.25 U/µl T7 RNA polymerase (50U/µl stock), 0.5 µg of DNA template and 1 µl of the [α-^{32}P]UTP (3000Ci/mmol, 10µCi/ml) in 20 µl final volume. The reaction was incubated at 37°C for 2 hours, after which 5 U of DNAse was added and incubated for another 30 min at 37°C. Then, the reaction was purified through nick-columns (Section 4.17.2). RNAs were separated on denaturing polyacrylamide gel (Section 4.16.3) to check quality and size. To generate probe for the RNAse protection assay, cold UTP were removed from the reaction and final concentration of 3NTP mix (ATP, CTP and GTP) was 0.4 mM (stock 4 mM). The probe was gel purified (Section 4.17.1).

4.16 Electrophoretic separation of RNA

4.16.1 Agarose gel

Agarose gel electrophoresis was used for checking the quality of in vitro transcribed RNA and total RNA after extraction from the cells and tissues. Agarose gels were pre-

pared as described for the DNA separation (Section 4.6), using the reagents and apparatus that were RNase free. 1% agarose gels were used for separation of total RNA, whereas 2% were used for in vitro transcribed RNA. The integrity of the total RNA extracted from cells and tissues was evaluated by the presence of three rRNA (28S, 18S and 5S).

4.16.2 Denaturing agarose gel

Denaturing agarose gels were used for separation of total RNAs prior to Northern transfer. 1.4% (w/v) agarose was melted in ddH_20 and put on a magnetic stirrer plate. 1X MOPS (from the 10X MOPS stock) was immediately added to the melted agarose. After 1-2 min 6% (v/v) formaldehyde (37% formaldehyde stock) was added, everything was mixed gently and immediately poured into the electrophoretic chamber (10x12cm) and left to solidify. The running buffer contained 1X MOPS and 3.7% (v/v) formaldehyde in ddH_20. RNA samples were resuspended in 1X RNA sample buffer and denatured at 65°C for 3 min prior to loading. The gels were usually run at 75 V for 4-5 hours. The presence of the three rRNA (28S, 18S and 5S) was visualized with the UV transiluminator.

4.16.3 Denaturing polyacrylamide gel

In vitro transcribed radioactive RNAs (hot RNAs) were run in denaturing polyacrylamide gels in order to check for their size and for the absence of extra bands that could appear due to incomplete transcription or degradation. Denaturing polyacrylamide gels were also used for hot RNA gel purification purposes and for the separation of RNA-protected fragments (RNase protection assay). 8M urea was dissolved in 6-10% (v/v) 19:1 acrylamide:bis-acrylamide (40% stock; AccuGel, National Diagnostics), 1X TBE and ddH_20 was added to the final volume. To allow the polymerization of the gel, 100 µl of a 10% (w/v) ammonium persulphate (APS) and 10 µl of TEMED were added for a 15 ml final volume and the mix was immediately poured into 0.75 mm tick, 10x12cm sized vertical electrophoretic apparatus. The gel was pre-run for about 1 hour at 200 V

in 1X TBE running buffer. RNA samples were resuspended in formamide-loading buffer and denatured at 95°C for 2 min prior to loading. The gel was usually run at 150-300 V, depending on the acrylamide concentration. 1 Kb DNA ladder (Invitrogen cat.No. 15615-024) was radioactively labeled and run in parallel in order to determine the size of the bands. Radioactive labeling was performed according to the manufacturer's T4 DNA Polymerase Labeling Protocol, with the exception that the [α-^{32}P]dCTP was used instead of [α-^{32}P]dATP. Additionally, the correct size of RNAs was estimated by the positions of bromophenol blue and xylene-cyanol that are present in the loading buffer. Migration of these two colors in the gel corresponds to the migration of RNA of specific size for each percentage of the gel. For example, in a 6% polyacrylamide gels migration of the bromophenol blue and xylene-cyanol coincide with the migration of 26 nt and 106 nt long RNAs, respectively. After the running, the gel was washed/fixed with fixer solution (10% methanol, 10% acetic acid) in order to eliminate urea allowing the adequate drying of the gel. The gel was dried at 80°C for 2 hours and exposed on a film (Kodak) or cyclone phospho-screen (Packard Bioscience Co.).

4.17 RNA purification

4.17.1 Gel purification

The radioactive RNA probe used for RNase protection assays was gel purified as follows. The hot RNA was *in vitro* synthesized as described in Section 4.15 and the whole reaction was run through a 6% denaturing polyacrylamide gel (4.16.3). The wet gel, covered with a thin plastic film, was exposed for 2 min and the borders of the gel were marked on the film. The film was developed and, with the help of the marks, it was precisely placed on the gel to cut-out the radioactive signal on the film corresponding to the desired RNA. The gel slice was transferred into a 1.5 ml tube containing 400 µl of

elution buffer (0.2% SDS, 1 mM EDTA pH 8.0, 0.5 M ammonium acetate) and left ON at RT°C with agitation. Next, the elution containing hot RNA was transferred into a clean 1.5 ml tube, acid phenol-chloroform purified and ethanol precipitated. Finally, the hot RNA was resuspended in R-loop buffer and used as a probe in the RNase protection assays.

4.17.2 Nick column purification

Nick column purification was used in order to purify *in vitro* transcribed hot RNAs (4.15) from the unincorporated radioactive nucleotides, which were subsequently used for EMSA experiments. The *in vitro* transcription reaction was passed through Illustra nick columns (GE Healthcare, cat.no. 17-0855-01). First, the original buffer inside the column was replaced by 3 ml of TE buffer. Then, the column was placed in a clean 1.5 ml tube and the *in vitro* transcription reaction was added. 350 µl TE buffer were added immediately and left to flow through. Next, the column was placed in a new clean 1.5 ml tube and 50 µl TE was added and left to flow through. This procedure was repeated for 7 times. Radioactivity was monitored on a geiger counter (Canberra Packard SRL) for each of the tubes and the elutions containing high picks of radioactivity were collected, ethanol precipitated and resuspended in ddH_20. The integrity of the RNAs was checked by running a denaturing polyacrylamide gel and subsequent exprosure on film.

4.17.3 Acid phenol-chloroform purification

Acid phenol-chloroform purification was used to purify cold *in vitro* transcribed RNAs and also was used after treatments with DNase or RNase (RNase protection assay). One volume of acid phenol-chloroform was added to the RNA sample, which is then vortexed and centrifuged for 5 min at 13200 rpm at 4°C. The supernatant is transferred into a clean 1.5 tube containing 2 volumes of 100% ethanol with 1/10 (v/v) of NaAc (3 M, pH 5.0). The sample is vortexed and left in dry ice for 20 min. Subsequently, the sample was centrifuged at 13200 rpm for 15 min at 4°C. The supernatant was removed; the pellet was washed with 500 µl of 70% ethanol and resuspended in ddH_20.

4.18 cDNA preparation

cDNAs were prepared by reverse transcription (RT) from total RNA. Briefly, 0.5-1 µg of total RNA was diluted to a final volume of 10 µl with ddH_20 and denatured at 70°C for 3 min. 10 µl of 2X reaction mix were added to the denatured RNA sample and the reaction was incubated for 30 min at 37°C. 10 µl of 2X reaction mix contained: 2.5 µl ddH_20, 4 µl 5X first strand buffer (Invitrogen), 1.25 µl DTT (stock 100 mM), 1.25 µl dNTPs (stock 5 mM), 0.75 µl random primer or oligo dT primer (stock 100 ng/µl) and 0.25 µl M-MLV Reverse Transcriptase (stock 200 U/µl, Invitrogen).

4.19 DNase treatment

DNase treatment of total RNA was used in order to eliminate transfected plasmid or genomic DNA contamination. 5 U of RNAse free-DNase (stock 1U/µl, Promega) were used for 5 µg of total RNA in a reaction that also contained 1X DNase buffer (stock 5X, Promega) and ddH_20 to a final volume. The reaction was then acid phenol-cloroform purified, ethanol precipitated and resuspended in ddH_20 or R-loop buffer, according to the experimental needs.

4.20 3'RACE

3'Rapid Amplification of cDNA Ends (3'RACE) was used to identify the 3'ends of the transcripts derived from the **A1-A23-A4[**ΔHm**]** and **A1-A23-A4[**ΔDSE**]** constructs (Figure 17). First, cDNA was synthesized from 1 µg of total RNA derived from the HeLa cells transfected with **A1-A23-A4[**ΔHm**]** and **A1-A23-A4[**ΔDSE**]** constructs. Then, two PCR

reactions were performed using the cDNA as template for the first PCR. The first PCR was performed using forward primer which annealed to the region ~500 nt upstream of the PAS4 and reverse primer was an oligo dT terminated in **A, C** or **G** linked to a specific 5'extension sequence (anchor). The obtained PCR products were used as template for the second PCR. The forward primer for the second PCR annealed to the region ~300 nt upstream of the PAS4 and the reverse primer was the sequence used as anchor in the first PCR. The products from the second PCR were gel purified, blunt-ended, cloned into pUC19 and fully sequenced. The primer pairs used for both PCRs are listed below.

First PCR:
Brain dir – 5' CCGGATCCTATAAGGTACCAGTGTCTCGTAAGTAATGCT 3'
dT Anchor 1 rev – 5' TAGGAATTCTCGAGCGGCCGC$(T)_{17}$ 3'

Second PCR:
mXhoI dir 1 – 5' CCGCTCGAGTCTGCCTTTCTCTATGCTAA 3'
Anchor 1 rev – 5' TAGGAATTCTCGAGCGGCCGC 3'

4.21 Quantitative PCR

Quantitative PCR (qPCR) was used to determine relative pre-mRNA (Figure 24B) or mRNA (Figure 36C and 38C) levels. A final volume of the PCR reaction was 20 µl and the mix for one sample contained: 10 µl 2X Syber green mix (BioRad, cat. No. 172-5006/25), 0.66 µl of each primer (stock 100 ng/µl) and ddH_20 to the final volume. Syber green mix contains DNA polymerase, dNTPs, reaction buffer and syber green fluorescent dye. 2 µl of cDNA were used as a template. The cDNA was prepared using total RNA previously treated with DNAse. The reactions were performed in duplicates for each sample. A Bio-Rad C1000 thermo-cycler coupled with a CFX96 real-time system was used for the amplification. The amplification reaction steps were as follows: a single denaturation step at 98°C for 30 sec, then 40 cycles of denaturation steps at 95°C for 5 sec, anneal-

ing/elongation steps at 58°C for 25 sec. In order to generate the amplicon's melting curves, the melting curve protocol followed the amplification reaction and comprised of one denaturation step at 95°C for 10 sec and and increase in the temperature from 65°C to 95°C at 0.5°C intervals, measuring the fluorescence at each step for 1 sec. The ΔΔCt between the amplicon of interest and reference amplicon was calculated using the Bio-Rad software. Primer pairs used for the qPCR reactions showed in the Figure 24B, 36C and 38C are listed below, respectively.

Alpha globin:
α Glo Ex2 real DIR – 5' TGGACGACATGCCCAAC 3'
α Glo Int2 real REV – 5' AACCCGCGTGATCCTCT 3'

GFP:
GFP 1004 FW – 5' TCAAGGAGGACGGCAACATC 3'
GFP 1124 RV – 5' TTGTGGCGGATCTTGAAGTTC 3'

Firefly luciferase:
LucqPCR 1580 dir - 5' TCGTTGACCGCCTGAAGTCT 3'
LucqPCR 1769 rev - 5' GGCGACGTAATCCACGATCT 3'

Renilla luciferase:
Renq PCR 1437 dir - 5' TGGAGCCATTCAAGGAGAAG 3'
Renq PCR 1649 rev - 5' TTCACGAACTCGGTGTTAGG 3'

Alpha globin:
agloEX1 real DIR - 5' GCCGACAAGACCAACGTCAA 3'
aGloex2 Real REV - 5' AGGTCGAAGTGCGGGAAGTA 3'

Renilla luciferase containing intron:
qREN1436D - 5' CTGGAGCCATTCAAGGAGAA 3'
qREN1554R - 5' GCGTTGTAGTTGCGGACAAT 3'

4.22 Northern blot analysis

In general, 5 µg of total RNA prepared from transfected cells were run on a denaturing agarose gel, as described in Section 4.16.2. The gel was visualized on a UV transilluminator and photographed. Then, the gel was washed three times for 15 min in ddH_20 and once in 20X SSC for 15 min. At the same time, a nitrocellulose membrane (Hybond-N^+, Amersham Pharmacia biotech) was incubated with 20X SSC for 15 min. Then, the RNA was transferred ON to the membrane by capillarity. The gel and the membrane were then visualized on the UV transilluminator and photographed, in order to check for the efficiency of the transfer. The positions of the 28S and 18S rRNAs were marked on the membrane with a pencil. The membrane was dried for 5 min at 37°C and the RNA was UV cross-linked, twice with 100x1200µJ (Stratalinker 1800). Then, the membrane was wet in 2X SSC and pre-hybridized for 1 hour at 42°C or 65°C (depending on the probe) in 5 ml of a commercial hybridization solution (Ultrasensitive Hybridisation Solution, Ambion) diluted with 2.5 ml of ddH_20. Next, the radioactive probe was denatured, added to the membrane and left to hybridize ON at 42°C or 65°C. The probe was then removed and the membrane washed using solutions containing descending concentrations of SSC (2X SSC, 1X SSC, 0.5X SSC and 0.1X SSC) and 0.1% SDS. Washing steps lasted 15 min for each solution at the temperature used for the hybridization. Radioactivity was checked after each washing with a geiger counter (Canberra Packard SRL) and when the background signal was sufficiently low, the membrane was exposed to a cyclone phosho-screen (Packard Bioscience Co.). The exposure time varied from 5 min to ON, depending on the signal strength. The signals were detected using cyclone reader (Packard Bioscience Co.) and the intensity of the obtained signals was quantified using OptiQuant that is part of cyclone software system. For the reutilization, membranes were exposed to boiling 0.5% SDS and incubated for 1 hour at 65°C, in order to remove the specific signal.

Radioactive probes were prepared using RediprimeII DNA labeling System (GE

Healthcare, cat.no.RPN1633). 50 ng of DNA in TE buffer were denatured at 95°C for 5 min. Then, denatured DNA was transferred into a Rediprime tube containing the reaction mix and 2 µl [α-^{32}P]dCTP (6000Ci/mmol, 10µCi/ml) was added. The reaction was incubated for 20 min at 37°C and subsequently purified from the unincorporated isotopes using Illustra nick column (GE Healthcare, cat.no. 17-0855-01). The nick column purification was performed as described in Section 4.17.2., with the exception that the final eluted probe was not precipitated, instead it was denatured and directly used for the hybridization.
The alpha globin probe was amplified from cDNA prepared from total RNA of HeLa cells transfected with the construct containing human alpha globin gene, and cloned into a pUC19 (from Dr. Laura de Conti, Molecular Pathology Group, ICGEB). The fragment is 272 nt long and contains the alpha globin exons 1 and 2. The alpha globin probe was cut from the pUC19 plasmid with BamHI and EcoRI and gel purified. The temperature used for the hybridization was 42°C.
The GFP probe (~500 nt long) was generated by a PCR reaction using the pEGFP-C2 plasmid as template. The PCR product was gel purified. The temperature used for the hybridization was 42°C. The primer pair used for the amplification were:
GFP813FW - 5' CGGCGTGCAGTGCTTCAGCCGCTAC 3'
GFPendRV - 5' CTTGTACAGCTCGTCCATGCCGAGAG 3'

The c-myc probe (~250 nt long) was generated by a PCR reaction using as template cDNA prepared from total RNA extracted from HeLa cells. The PCR product was gel purified. The temperature used for the hybridization was 65°C. The primer pair used for the amplification were:
c-myc ex 2 dir - 5' CCTACCCTCTCAACGACAGC 3'
c-myc ex 3 rev - 5' CTCTGACCTTTTGCCAGGAG 3'

The GAPDH probe (~1.2 Kb long) was amplified from cDNA prepared from total RNA of HeLa cells and cloned into a pUC19. The GAPDH probe was cut from the plasmid with PstI and gel purified. The temperature used for the hybridization was 42°C.

4.23 RNase protection assay

The RNase protection assay was used to determine the cleavage efficiency of the transcripts derived from HeLa cells transfected with the different minigene constructs (Figure 23). The RNase protection assays were performed as described in Sambrook et al 1989. and Dalziel et al 2007.

Briefly, the radioactively labeled probe was added to the RNA sample in a final volume of 30 µl in R-loop buffer. Thus prepared hybridization mix was denatured at 95°C for 10 min and left to hybridize at 45°C for 15-18 hours in a water bath. Then, 300 µl of RNase mix were added to the hybridized sample and the digestion was carried out for 45 min at 30°C in a water bath. The RNase mix was freshly prepared and contained 10 mM Tris-HCl pH 7.4, 300 mM NaCl, 5 mM EDTA, 0.5U/µl RNase T1 and 0.01 mg/µl RNase A. After the RNase digestion, the sample was incubated with 100 µl Proteinase K mix for 15 min at 30°C in a water bath. 100 µl Proteinase K mix contained 96 µl Proteinase K buffer (100 mM Tris-HCl pH7.5, 2% SDS, EDTA 20 mM) and 4 µl Proteinase K (stock 10 mg/ml). Then, the sample was acid-phenol purified, ethanol precipitated and resuspended in 10 µl formamide-loading buffer. Half of the resuspension was denatured for 2 min at 95°C and loaded onto a 6% polyacrylamide 8M urea gel (see Section 4.16.3). In parallel, 10 cps of probe and radioactively labeled 1 Kb marker were run and the gel was washed with fixer solution, dried at 80°C for 2 hours and exposed to a cyclone phospho-screen. The signals were visualized on a Cyclone imager (Packard Bioscience Co.) and quantified using OptiQuant.

RNA samples were prepared from total RNA of transfected HeLa cells. Total RNA was treated with DNase, acid phenol-chloroform purified, ethanol precipitated and resuspended in R-loop buffer. The quality of the RNA was checked by running 1% agarose gels and the concentration was measured by a "nanodrop" spectrophotometer. The quantity of each RNA sample used for the RNase protection assay was normalized for the transfection efficiency. 400 ng of one RNA sample were used for the assay and this sample was

chosen as reference in the normalization calculations. 10 µg of mouse brain total RNA were used as a control for the correct cleavage protected fragment. Total RNA from mouse brain was prepared using TriFast reagent and resuspended in R-loop buffer.
For the RNA probe preparation, the fragment upstream and downstream of the PAS4 (Figure 23B) was amplified using following primer pair:

mXhoI dir 2 - 5' CCGCTCGAGATTTATGGGTTGGTTTA 3'
mADD2 BR short rev - 5' CGGGATCCCCGCGGGAATTCCCTACATTTC 3'

The PCR product was cloned into a pUC19 plasmid, sequenced and then cloned in the antisense orientation downstream of the T7 promoter in the PBS II KS plasmid (Stratagene) in the SacII and XhoI sites. The plasmid was linearized with XhoI, phenol-chloroform purified, ethanol precipitated and used for the *in vitro* transcription reaction (for the conditions see Section 4.15). The *in vitro* transcribed probe was gel purified through 6% polyacrylamide 8M urea gel (see Section 4.17.1) and resuspended in R-loop buffer. Approximately 500 cps of the probe were used for the hybridization reactions.

4.24 EMSA

Electrophoretic Mobility Shift Assay (EMSA) were performed to detect RNA-protein interactions (Figure 33 and 34). The RNA-protein reaction mix contained: ~150-200 cps of hot RNA, 1X binding buffer (stock 10X), 2-3 µg/µl HeLa nuclear extract (Cilbiotech, stock 10-12 µg/µl), 10 µg/µl heparin (stock 200 µg/µl) and ddH2O to the final volume. 10X binding buffer contained 52 mM Hepes pH 7.5, 10 mM MgCl2, 8 mM Mg acetate, 5.2 mM DTT, 7.5 mM ATP, 10 mM GTP, 38% glycerol and ddH20 to the final volume. First, nuclear protein extract was mixed with heparin and left for 5-10 min at RT. In the meantime, the hot RNA, binding buffer and ddH20 were mixed. Nuclear extract and heparin were then added to the mix and everything was incubated for 20 min at RT. After 20 min, the mix was loaded onto a 6% native polyacrylamide gel and run at 200 V at 4°C. One well was loaded with 5 µl of loading buffer (30% glycerol, 0,25% bromophenol blue) and the

gel was run until the blue reached two thirds of the gel. The gels contained: 6% 37.5:1 acrylamide:bis-acrylamide (30% stock, Promega, Serva products), 0.5X TBE, ddH20, 140µl of a 10% (w/v) ammonium persulphate (APS) and 17.5 µl of TEMED were added for a 17.5 ml final volume. 0.5X TBE was used as running buffer. The gel was pre-run for 1 hour at 4°C. After the running, the gel was dried at 80°C for 2 hours, exposed on cyclone phosphor-screen and visualized on a cyclone imager (Packard Bioscience Co.). Hot RNAs were in vitro transcribed (as described in Section 4.15) using as a template the PCR products previously purified using the PCR clean-up protocol (see Section 4.5.2). The primer pairs used for the amplification of the different templates for the *in vitro* transcription are listed below. The forward primer contained the T7 promoter sequence and is represented in lower case.

WT

T7pull244UPdir – 5' tacgtaatacgactcactataggCTCGTTCCTGAAGAAA 3'
pull203rev – 5' AGCAGGACCCACAGAGCCACT 3'

ΔUP

T7pull268UPdir – 5' tacgtaatacgactcactataggGGGAAGGAGAAGGTGGAA 3'
pull203rev – 5' AGCAGGACCCACAGAGCCACT 3'

ΔDOWN

T7pull244UPdir – 5' tacgtaatacgactcactataggCTCGTTCCTGAAGAAA 3'
pullSTOPrev – 5' TCAGGATTCCACCTTCTCCTT 3'

In vitro transcribed RNAs were purified through nick columns and their quality was checked by running 10% polyacrylamide 8M urea gels.

For the EMSA competition assays, cold competitor RNAs were in vitro transcribed (see Section 4.15) using the same templates as described above, acid phenol-chloroform purified, ethanol precipitated and resuspended in ddH20. The quality of the RNAs was checked by running 2% denaturing agarose gel.

4.25 Pull down analysis

Pull down analysis were performed in order to detect single protein differences between the proteins bound to the different RNAs (Figure 35). The cold RNAs used for the pull down analysis were the same ones used for the EMSA competition analysis.

Approximately, 500 pmoles of RNA were incubated with 5 mM sodium m-periodate (Sigma) solution in 0.1 M NaOAc (pH 5.0), in a final volume of 400 µl at RT for 1 hour in the dark with agitation. Then, the RNA was ethanol-precipitated and resuspended in 100 µl of 0.1 M NaOAc (pH 5.0). In the meantime, 500 µl of adipic acid dehydrazide agarose bead 50% slurry (Sigma) were washed four times with 10 ml 0.1 M NaOAc (pH 5.0). After the final washing, the beads were resuspended in 300 µl of 0.1 M NaOAc (pH 5.0) and added to the previously prepared 100 µl RNA. The mix was incubated ON in the dark at 4°C with agitation. Then, the beads with the bound RNA were centrifuged at 4000 rpm for 5 min, and the supernatant was discarded. The RNA-loaded beads were washed twice with 2 M NaCl and three times with washing buffer (52 mM Hepes pH 7.5, 10 mM MgCl2, 8 mM Mg acetate, 5.2 mM DTT, 38% glycerol and ddH20 to a final volume). Subsequently, the RNA loaded beads were incubated with 20 µg/µl HeLa nuclear extract, 1X binding buffer (the same one used for the EMSA experiments) and ddH20 to a 500 µl final volume, for 20 min at RT with agitation. Then, the reaction was centrifuged at 4000 rpm for 3 min, the supernatant discarded, and washed four times with washing buffer. The final pellet was resuspended with 60 µl of SDS loading buffer (20% SDS, 1 M DTT, 0.63 M Tris-HCl pH 7.0, 0.2% bromophenol blue, 20% glycerol, 10 mM EDTA pH 7.0), denatured at 95°C for 5 min and loaded onto a 10% SDS-PAGE gel. For a final volume of 20 ml, the running gel contained: 6.7 ml of 37.5:1 acrylamide:bis-acrylamide (30% stock, Promega, Serva products), 5 ml Tris-HCl (stock 1.5M, pH 8.8), 8.3 ml ddH20, 200 µl 10% SDS, 200 µl 10% APS and 20 µl TEMED. The stacking gel contained: 0.56 ml of 37.5:1 acrylamide:bis-acrylamide (30% stock, Promega, Serva products), 2.5 ml Tris-HCl (stock 0.5 M, pH 6.8), 1.86 ml ddH20, 50 µl 10% SDS, 50 µl APS and 10 µl

TEMED. The gel was run at 25-30 mA for 4-5 hours, or ON at 5 mA, in 1X SDS-PAGE running buffer (50 mM Tris, 0.38 M glycine, 0.1% SDS). After running, the gels were stained with colloidal coomassie. The bands showing single protein differences (Figure 34) were excised, digested with trypsin and identified by mass spectrometry analysis. The peptide sequence analysis was performed using an electrospray ionization mass spectrometer (LCQ DECA XP, ThermoFinnigam). The mass spectrometry analysis was performed by the Protein Neworks Group at the ICGEB.

4.26 Luciferase reporter assay

The luciferase reporter assays were performed according to the Promega Dual-Luciferase Reporter Assay System instructions. All solutions were purchased from Promega.
Cells were transfected with the reporter constructs in triplicates, as described in Section 4.13. 24 hours after transfection, medium was removed, cells were washed twice with PBS and 100 µl of Passive Lysis Solution were added. The cells were incubated for 15 min at RT with agitation. Then, the lysates were transferred into 1.5 ml tubes and centrifuged for 4 min at 13200 rpm at 4°C. 10 µl of the supernatant were transferred into a clean 1.5 ml tube and used to determine luciferase activity with a Turner Biosystems Luminometer 20/20. First, 50 µl of LARII solution were added, the reaction was mixed by pipetting and the luminescent coming from the firefly luciferase activity was measured. Next, 50 µl of Stop&Glow solution was added in order to to stop the activity of the firefly luciferase and to measure the luminescent coming from renilla luciferase activity. The ratio between firefly and renilla luciferase activity was defined as relative luciferase activity.

REFERENCES

Amara, S.G., Evans, R.M., and Rosenfeld, M.G. (1984). Calcitonin/calcitonin gene-related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. Mol Cell Biol 4, 2151-2160.

Antoniou, M., Geraghty, F., Hurst, J., and Grosveld, F. (1998). Efficient 3'-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. Nucleic Acids Res 26, 721-729.

Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C., and Wilusz, J. (2002). Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. Nucleic Acids Res 30, 1842-1850.

Ayala, Y.M., De Conti, L., Avendano-Vazquez, S.E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E., et al. (2011). TDP-43 regulates its mRNA levels through a negative feedback loop. EMBO J 30, 277-288.

Bagga, P.S., Arhin, G.K., and Wilusz, J. (1998). DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. Nucleic Acids Res 26, 5343-5350.

Bak, M., Silahtaroglu, A., Moller, M., Christensen, M., Rath, M.F., Skryabin, B., Tommerup, N., and Kauppinen, S. (2008). MicroRNA expression in the adult mouse central nervous system. RNA 14, 432-444.

Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev 11, 1703-1716.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. Genome Res 10, 1001-1010.

Bednarek, E., and Caroni, P. (2011). beta-Adducin is required for stable assembly of new synapses and improved memory upon environmental enrichment. Neuron 69, 1132-1146.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72, 291-336.

Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25, 106-110.

Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, I.W., and Gunderson, S.I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell 72, 881-892.

Brackenridge, S., and Proudfoot, N.J. (2000). Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. Mol Cell Biol 20, 2660-2669.

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol 16, 107-113.

Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol Cell 12, 1467-1476.

Brusa, R., Zimmermann, F., Koh, D.S., Feldmeyer, D., Gass, P., Seeburg, P.H., and Sprengel, R. (1995). Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. Science 270, 1677-1680.

Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. Curr Opin Cell Biol 17, 257-261.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdiscip Rev RNA 3, 1-12.

Caputi, M., Casari, G., Guenzi, S., Tagliabue, R., Sidoli, A., Melo, C.A., and Baralle, F.E. (1994). A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. Nucleic Acids Res 22, 1018-1022.

Chen, F., MacDonald, C.C., and Wilusz, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res 23, 2614-2620.

Chen, J.M., Ferec, C., and Cooper, D.N. (2006). A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet 120, 1-21.

Costessi, L., Devescovi, G., Baralle, F.E., and Muro, A.F. (2006). Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. Nucleic Acids Res 34, 243-253.

Cusi, D., Barlassina, C., Azzani, T., Casari, G., Citterio, L., Devoto, M., Glorioso, N., Lanzani, C., Manunta, P., Righetti, M., et al. (1997). Polymorphisms of alpha-adducin and salt sensitivity in patients with essential hypertension. Lancet 349, 1353-1357.

Dai, W., Zhang, G., and Makeyev, E.V. (2012). RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. Nucleic Acids Res 40, 787-800.

Dalziel, M., Nunes, N.M., and Furger, A. (2007). Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. Mol Cell Biol 27, 1568-1580.

Dani, C., Blanchard, J.M., Piechaczyk, M., El Sabouty, S., Marty, L., and Jeanteur, P. (1984). Extreme instability of myc mRNA in normal and transformed human cells. Proc Natl Acad Sci U S A 81, 7046-7050.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature 389, 399-402.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., and Barabino, S.M. (2004). Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. J Biol Chem 279, 35788-35797.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and consequences of alternative polyadenylation. Mol Cell 43, 853-866.

Dominski, Z., and Marzluff, W.F. (2007). Formation of the 3' end of histone mRNA: getting closer to the end. Gene 396, 373-390.

Dugas, J.C., Cuellar, T.L., Scholze, A., Ason, B., Ibrahim, A., Emery, B., Zamanian, J.L., Foo, L.C., McManus, M.T., and Barres, B.A. (2010). Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. Neuron 65, 597-611.

Dye, M.J., and Proudfoot, N.J. (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol Cell 3, 371-378.

Dye, M.J., and Proudfoot, N.J. (2001). Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. Cell 105, 669-681.

Edbauer, D., Neilson, J.R., Foster, K.A., Wang, C.F., Seeburg, D.P., Batterton, M.N., Tada, T., Dolan, B.M., Sharp, P.A., and Sheng, M. (2010). Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. Neuron 65, 373-384.

Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A 94, 11893-11898.

Fogel, B.L., McNally, L.M., and McNally, M.T. (2002). Efficient polyadenylation of Rous sarcoma virus RNA requires the negative regulator of splicing element. Nucleic

Acids Res 30, 810-817.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19, 92-105.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol 8, 113-126.

Gilligan, D.M., Lozovatsky, L., Gwynn, B., Brugnara, C., Mohandas, N., and Peters, L.L. (1999). Targeted disruption of the beta adducin gene (Add2) causes red blood cell spherocytosis in mice. Proc Natl Acad Sci U S A 96, 10717-10722.

Gilmartin, G.M., and Nevins, J.R. (1991). Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol Cell Biol 11, 2432-2438.

Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K., et al. (2011). Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. PLoS Genet 7, e1002390.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol Cell Biol 26, 3986-3996.

Guan, F., Caratozzolo, R.M., Goraczniak, R., Ho, E.S., and Gunderson, S.I. (2007). A bipartite U1 site represses U1A expression by synergizing with PIE to inhibit nuclear polyadenylation. RNA 13, 2129-2140.

Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C., and Mattaj, L.W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell 76, 531-541.

Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol Cell 1, 255-264.

Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I.W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. Genes Dev 11, 761-773.

Gurevich, I., Tamir, H., Arango, V., Dwork, A.J., Mann, J.J., and Schmauss, C. (2002). Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. Neuron 34, 349-356.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem J 430, 379-392.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J 30, 4414-4422.

Hedley, M.L., and Maniatis, T. (1991). Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. Cell 65, 579-586.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H. (2001). Quality control of mRNA 3'-end processing is linked to the nuclear exosome. Nature 413, 538-542.

Huang, Y.S., Carson, J.H., Barbarese, E., and Richter, J.D. (2003). Facilitation of dendritic mRNA transport by CPEB. Genes Dev 17, 638-653.

Hwang, H.W., Wentzel, E.A., and Mendell, J.T. (2007). A hexanucleotide element directs microRNA nuclear import. Science 315, 97-100.

Janas, M.M., Khaled, M., Schubert, S., Bernstein, J.G., Golan, D., Veguilla, R.A., Fisher, D.E., Shomron, N., Levy, C., and Novina, C.D. (2011). Feed-forward microprocessing and splicing activities at a microRNA-containing intron. PLoS Genet 7, e1002330.

Jeffries, C.D., Fried, H.M., and Perkins, D.O. (2011). Nuclear and cytoplasmic localization of neural stem cell microRNAs. RNA 17, 675-686.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kuhn, U., Menzies, F.M., Oude Vrielink, J.A., Bos, A.J., Drost, J., Rooijers, K., et al. (2012a). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149, 538-553.

Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A 106, 7028-7033.

Ji, Z., and Tian, B. (2009). Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PLoS One 4, e8419.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12, 5-14.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468, 664-668.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J 23, 616-626.

Keegan, L.P., Gallo, A., and O'Connell, M.A. (2001). The many roles of an RNA editor.

Nat Rev Genet 2, 869-878.

Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. (1991). Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. EMBO J 10, 4241-4249.

Kim, D.H., Saetrom, P., Snove, O., Jr., and Rossi, J.J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. Proc Natl Acad Sci U S A 105, 16230-16235.

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., and Handa, H. (2006). Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. Nucleic Acids Res 34, 6264-6271.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Mol Cell 23, 195-205.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401-1414.

Lavigueur, A., La Branche, H., Kornblihtt, A.R., and Chabot, B. (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev 7, 2405-2417.

Legendre, M., and Gautheret, D. (2003). Sequence determinants in human polyadenylation site selection. BMC Genomics 4, 7.

Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur J Biochem 247, 461-469.

Liao, J.Y., Ma, L.M., Guo, Y.H., Zhang, Y.C., Zhou, H., Shao, P., Chen, Y.Q., and Qu, L.H. (2010). Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. PLoS One 5, e10563.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464-469.

Lin, C.L., Evans, V., Shen, S., Xing, Y., and Richter, J.D. (2010). The nuclear experience of CPEB: implications for RNA processing and translational control. RNA 16, 338-348.

Lou, H., Helfman, D.M., Gagel, R.F., and Berget, S.M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. Mol Cell Biol 19, 78-85.

Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Mol Cell Biol 18, 4977-4985.

Lou, H., Yang, Y., Cote, G.J., Berget, S.M., and Gagel, R.F. (1995). An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. Mol Cell Biol 15, 7135-7142.

Lutz, C.S., and Alwine, J.C. (1994). Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. Genes Dev 8, 576-586.

Lutz, C.S., and Moreira, A. (2011). Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip Rev RNA 2, 22-31.

Lutz, C.S., Murthy, K.G., Schek, N., O'Connor, J.P., Manley, J.L., and Alwine, J.C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev 10, 325-337.

Maas, S., Patt, S., Schrey, M., and Rich, A. (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proc Natl Acad Sci U S A 98, 14687-14692.

MacDonald, C.C., Wilusz, J., and Shenk, T. (1994). The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol 14, 6647-6654.

Maciolek, N.L., and McNally, M.T. (2007). Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus. J Virol 81, 11208-11217.

Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell 27, 435-448.

Malumbres, M. (2012). miRNAs and cancer: An epigenetics view. Mol Aspects Med

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes Dev 24, 1073-1074.

Mansfield, K.D., and Keene, J.D. (2012). Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res 40, 2734-2746.

Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. (2010). Crosstalk between mRNA 3' end processing and transcription initiation. Mol Cell 40, 410-422.

Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol 10, 1102-1109.

Martinson, H.G. (2011). An active role for splicing in 3'-end formation. Wiley Interdiscip Rev RNA 2, 459-470.

Matsuoka, Y., Li, X., and Bennett, V. (2000). Adducin: structure, function and regulation. Cell Mol Life Sci 57, 884-895.

Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138, 673-684.

McCracken, S., Lambermon, M., and Blencowe, B.J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. Mol Cell Biol 22, 148-160.

McCracken, S., Longman, D., Johnstone, I.L., Caceres, J.F., and Blencowe, B.J. (2003). An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation. J Biol Chem 278, 44153-44160.

Mendez, R., and Richter, J.D. (2001). Translational control by CPEB: a means to the end. Nat Rev Mol Cell Biol 2, 521-529.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. Cell.

Millevoi, S., Decorsiere, A., Loulergue, C., Iacovoni, J., Bernat, S., Antoniou, M., and Vagner, S. (2009). A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. Nucleic Acids Res 37, 4672-4683.

Millevoi, S., Geraghty, F., Idowu, B., Tam, J.L., Antoniou, M., and Vagner, S. (2002). A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. EMBO Rep 3, 869-874.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. EMBO J 25, 4854-4864.

Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L., and Proudfoot, N.J. (1998). The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes Dev 12, 2522-2534.

Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N.J. (1995). Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. EMBO J 14, 3809-3819.

Moreno-Morcillo, M., Minvielle-Sebastia, L., Mackereth, C., and Fribourg, S. (2011). Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure. RNA 17, 412-418.

Muro, A.F., Marro, M.L., Gajovic, S., Porro, F., Luzzatto, L., and Baralle, F.E. (2000). Mild spherocytic hereditary elliptocytosis and altered levels of alpha- and gamma-adducins in beta-adducin-deficient mice. Blood 95, 3978-3985.

Murthy, K.G., and Manley, J.L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev 9, 2672-2683.

Niwa, M., and Berget, S.M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev 5, 2086-2095.

Niwa, M., Rose, S.D., and Berget, S.M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev 4, 1552-1559.

Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. EMBO J 29, 1523-1536.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat Genet 36, 1014-1018.

Pandya-Jones, A. (2011). Pre-mRNA splicing during transcription in the mammalian system. Wiley Interdiscip Rev RNA 2, 700-717.

Pauws, E., van Kampen, A.H., van de Graaf, S.A., de Vijlder, J.J., and Ris-Stalpers, C. (2001). Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. Nucleic Acids Res 29, 1690-1694.

Perkins, K.J., Lusic, M., Mitar, I., Giacca, M., and Proudfoot, N.J. (2008). Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. Mol Cell 29, 56-68.

Pinto, P.A., Henriques, T., Freitas, M.O., Martins, T., Domingues, R.G., Wyrzykowska, P.S., Coelho, P.A., Carmo, A.M., Sunkel, C.E., Proudfoot, N.J., et al. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. EMBO J 30, 2431-2444.

Pique, M., Lopez, J.M., Foissac, S., Guigo, R., and Mendez, R. (2008). A combinatorial code for CPE-mediated translational control. Cell 132, 434-448.

Porro, F., Rosato-Siri, M., Leone, E., Costessi, L., Iaconcig, A., Tongiorgi, E., and Muro, A.F. (2010). beta-adducin (Add2) KO mice show synaptic plasticity, motor coordination and behavioral deficits accompanied by changes in the expression and phosphorylation levels of the alpha- and gamma-adducin subunits. Genes Brain Behav 9, 84-96.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol 16, 272-278.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev 25, 1770-1782.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211-214.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell 108, 501-512.

Richter, J.D. (2001). Think globally, translate locally: what mitotic spindles and neuronal synapses have in common. Proc Natl Acad Sci U S A 98, 7069-7071.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. RNA 15, 823-836.

Rosenstierne, M.W., Vinther, J., Mittler, G., Larsen, L., Mann, M., and Norrild, B. (2008). Conserved CPEs in the p53 3' untranslated region influence mRNA stability and protein synthesis. Anticancer Res 28, 2553-2559.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). Molecular Cloning-A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320, 1643-1647.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res 19, 381-394.

Schek, N., Cooke, C., and Alwine, J.C. (1992). Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. Mol Cell Biol 12, 5386-5393.

Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. Trends Biochem Sci 33, 501-510.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res 18, 5799-5805.

Shell, S.A., Hesse, C., Morris, S.M., Jr., and Milcarek, C. (2005). Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. J Biol Chem 280, 39950-39961.

Shepard, P.J., Choi, E.A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17, 761-772.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell 33, 365-376.

Shomron, N., and Levy, C. (2009). MicroRNA-biogenesis and Pre-mRNA splicing crosstalk. J Biomed Biotechnol 2009, 594678.

Soller, M., and White, K. (2003). ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA. Genes Dev 17, 2526-2538.

Sonenberg, N. (1988). Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation. Prog Nucleic Acid Res Mol Biol 35, 173-207.

Sullivan, K.D., Steiniger, M., and Marzluff, W.F. (2009). A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol Cell 34, 322-332.

Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell 87, 941-952.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33, 201-212.

Tripodi, G., Piscone, A., Borsani, G., Tisminetzky, S., Salardi, S., Sidoli, A., James, P., Pongor, S., Bianchi, G., and Baralle, F.E. (1991). Molecular cloning of an adducin-like protein: evidence of a polymorphism in the normotensive and hypertensive rats of the Milan strain. Biochem Biophys Res Commun 177, 939-947.

Vagner, S., Vagner, C., and Mattaj, I.W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev 14, 403-413.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev 19, 1315-1327.

Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J., and Milcarek, C. (2001). hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. Mol Cell Biol 21, 1228-1238.

Vessey, J.P., Schoderboeck, L., Gingl, E., Luzi, E., Riefler, J., Di Leva, F., Karra, D., Thomas, S., Kiebler, M.A., and Macchi, P. (2010). Mammalian Pumilio 2 regulates dendrite

morphogenesis and synaptic function. Proc Natl Acad Sci U S A 107, 3222-3227.

von Roretz, C., Di Marco, S., Mazroui, R., and Gallouzi, I.E. (2011). Turnover of AU-rich-containing mRNAs during stress: a matter of survival. Wiley Interdiscip Rev RNA 2, 336-347.

Vorlova, S., Rocco, G., Lefave, C.V., Jodelka, F.M., Hess, K., Hastings, M.L., Henke, E., and Cartegni, L. (2011). Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. Mol Cell 43, 927-939.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470-476.

Wang, Y., Fairley, J.A., and Roberts, S.G. (2010). Phosphorylation of TFIIB links transcription initiation and termination. Curr Biol 20, 548-553.

Weiss, E.A., Gilmartin, G.M., and Nevins, J.R. (1991). Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J 10, 215-219.

Wilusz, J.E., and Beemon, K.L. (2006). The negative regulator of splicing element of Rous sarcoma virus promotes polyadenylation. J Virol 80, 9634-9640.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434, 338-345.

Yang, Q., Gilmartin, G.M., and Doublie, S. (2010). Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. Proc Natl Acad Sci U S A 107, 10062-10067.

Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., and Hovorun, D.M. (2003). Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic Acids Res 31, 1375-1386.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev 63, 405-445.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep gratitude to my supervisor, Dr. Andrés F. Muro, for making a scientist out of me. Thank you, Andrés!

A big thank you goes also to my external supervisor, Dr. André Furger, most of all for his enthusiasm and positive energy that motivated me throughout my PhD studies.

It was a real pleasure and honor for me to do my PhD at the ICGEB and for this opportunity I am grateful and indebted to prof. F. E. Baralle. He is an extraordinary scientist and above all, a wonderful person.

I would like to thank to all the present and past members of the Mouse Molecular Genetics and Molecular Pathology groups, for their technical help, discussions and suggestions. I would also like to extend the thanks to all of my ICGEB family for creating the enjoyable atmosphere to work in.

I would also like to thank to Bob Marley, U2, Queen, Sting, Parni valjak and many others whose music was a big support for me during these years.

I found myself very fortunate, as during these years of the PhD, I have found friends for a lifetime. I am thankful to them for all the support, encouragement, discussions and unconditional help.

Grazie, mio amor!

И на крају бих желела да се захвалим мојим родитељима и брату, за сву њихову љубав, брижност и подршку. Њима посвећујем ову тезу.